

D. J.
#7 1-24-02
Priority Papers
337 E. 10th St. (2101)
1033 U.S. PTO
09/862437
05/23/01
Attorney Docket No. 826.1726

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of:

Yoshio NAKAO

Application No.:

Group Art Unit:

Filed: May 22, 2001

Examiner:

For: APPARATUS FOR READING A PLURALITY OF DOCUMENTS AND A METHOD
THEREOF

**SUBMISSION OF CERTIFIED COPY OF PRIOR FOREIGN
APPLICATION IN ACCORDANCE
WITH THE REQUIREMENTS OF 37 C.F.R. §1.55**

Assistant Commissioner for Patents
Washington, D.C. 20231

Sir:

In accordance with the provisions of 37 C.F.R. §1.55, the applicant(s) submit(s) herewith
a certified copy of the following foreign application:

Japanese Patent Application No. 2000-290886

Filed: September 25, 2000

It is respectfully requested that the applicant(s) be given the benefit of the foreign filing
date(s) as evidenced by the certified papers attached hereto, in accordance with the
requirements of 35 U.S.C. §119.

Date: *5/24/01*

Respectfully submitted,

STAAS & HALSEY LLP

By: *[Signature]*

James D. Halsey, Jr.
Registration No. 22,729

700 11th Street, N.W., Ste. 500
Washington, D.C. 20001
(202) 434-1500

©2001 Staas & Halsey LLP

PATENT OFFICE
JAPANESE GOVERNMENT



This is to certify that the annexed is a true copy of the following application as filed with this Office.

Date of Application: September 25, 2000

Application Number: Patent Application
No. 2000-290886

Applicant(s): FUJITSU LIMITED

December 22, 2000

Commissioner,
Patent Office Kozo Oikawa

Certificate No. 2000-3105880

日 本 国 特 許 庁
PATENT OFFICE
JAPANESE GOVERNMENT

11033 U.S. PTO
09/862437
05/23/01

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日
Date of Application: 2000年 9月25日

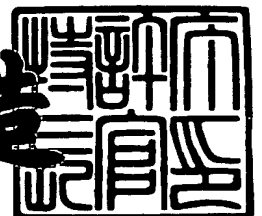
出 願 番 号
Application Number: 特願2000-290886

出 願 人
Applicant(s): 富士通株式会社

2000年12月22日

特許庁長官
Commissioner,
Patent Office

及 川 耕 造



出証番号 出証特2000-3105880

【書類名】 特許願

【整理番号】 0051731

【提出日】 平成12年 9月25日

【あて先】 特許庁長官殿

【国際特許分類】 G06F 17/20

【発明の名称】 複数の文書を閲覧するための装置および方法

【請求項の数】 9

【発明者】

 【住所又は居所】 神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内

 【氏名】 仲尾 由雄

【特許出願人】

 【識別番号】 000005223

 【氏名又は名称】 富士通株式会社

【代理人】

 【識別番号】 100074099

 【住所又は居所】 東京都千代田区二番町8番地20 二番町ビル3F

 【弁理士】

 【氏名又は名称】 大菅 義之

 【電話番号】 03-3238-0031

【選任した代理人】

 【識別番号】 100067987

 【住所又は居所】 神奈川県横浜市鶴見区北寺尾7-25-28-503

 【弁理士】

 【氏名又は名称】 久木元 彰

 【電話番号】 045-573-3683

【手数料の表示】

 【予納台帳番号】 012542

 【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9705047

【ブルーフの要否】 要

【書類名】 明細書

【発明の名称】 複数の文書を開覧するための装置および方法

【特許請求の範囲】

【請求項 1】 開覧対象として指定された複数の文書を利用者に提示する文書開覧装置であって、

前記複数の文書のそれぞれの話題階層を認定する話題階層認定手段と、

認定された話題階層に基づき、前記複数の文書に共通して現れる関連話題を抽出する関連話題抽出手段と、

抽出された関連話題に対応する記述箇所を、該複数の文書のそれぞれから切り出して出力する関連話題提示手段と

を備えることを特徴とする文書開覧装置。

【請求項 2】 前記関連話題抽出手段は、前記認定された話題階層中の様々な粒度の話題からなる話題対について、該話題対の各話題に対応する記述箇所の語彙的類似性に基づき、該話題対の話題間の関連度を求め、話題の包含関係に基づいて設定された閾値以上の関連度を持つ話題対を、前記関連話題として抽出することを特徴とする請求項 1 記載の文書開覧装置。

【請求項 3】 前記関連話題提示手段は、切り出された記述箇所を、前記複数の文書間の関連箇所として対置して提示することを特徴とする請求項 1 記載の文書開覧装置。

【請求項 4】 前記関連話題提示手段は、前記関連箇所と元の文書とを対置・連動して提示することを特徴とする請求項 3 記載の文書開覧装置。

【請求項 5】 前記関連話題提示手段は、前記関連箇所の内容を要約して提示することを特徴とする請求項 3 記載の文書開覧装置。

【請求項 6】 前記関連話題提示手段は、前記複数の文書に対応する複数の話題階層と、前記複数の関連話題に基づく、該複数の話題階層中の話題間の対応関係とを、図面を用いて提示し、該図面上における前記利用者の指示に応じて、該複数の文書中の対応箇所を連動して提示することを特徴とする請求項 3 記載の文書開覧装置。

【請求項 7】 前記関連話題提示手段は、前記複数の文書のうちの 1 つの文

書を基準文書として、前記関連話題に対応する記述箇所を他の文書から該基準文書に取り込んで新たな統合文書を作成し、該統合文書を出力することを特徴とする請求項 1 記載の文書閲覧装置。

【請求項 8】 閲覧対象として指定された複数の文書を利用者に提示するコンピュータのためのプログラムを記録した記録媒体であって、該プログラムは、前記複数の文書のそれぞれの話題階層を認定し、

認定された話題階層に基づき、前記複数の文書に共通して現れる関連話題を抽出し、

抽出された関連話題に対応する記述箇所を、該複数の文書のそれぞれから切り出して出力する

処理を前記コンピュータに実行させることを特徴とするコンピュータ読み取り可能な記録媒体。

【請求項 9】 閲覧対象として指定された複数の文書を利用者に提示する文書提示方法であって、

前記複数の文書のそれぞれの話題階層を認定し、

認定された話題階層に基づき、前記複数の文書に共通して現れる関連話題を抽出し、

抽出された関連話題に対応する記述箇所を、該複数の文書のそれぞれから切り出して出力する

ことを特徴とする文書提示方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、機械可読文書をコンピュータの画面上で閲覧するための装置およびその方法に関し、特に、複数文書の内容を対比して閲覧可能な形で提示することで、関連文書の比較閲覧作業を支援することを意図したものである。

【0002】

【従来の技術】

利用者が複数の関連文書を比較しながら閲覧する場合、それらの文書の関連箇

所をわかりやすく提示する技術があれば、比較作業の効率が向上する。例えば、ある調査項目について複数の地域の実情を調査レポートにまとめるために、各地域の調査担当者から寄せられた調査レポートを読む場合、あるいは、質問状と回答書を読み比べる場合等に、関連箇所の比較を支援する技術が要望される。このような複数文書の比較支援に関する文献として、例えば、以下の7つが挙げられる。

[1] Christine M. Neuwirth and David S. Kaufer. The role of external representations in the writing process: Implications for the design of hypertext-based writing tools. In Proc. of Hypertext '89, pp.319-341. the Association for Computing Machinery, Nov. 1989.

[2] 大森信行(Nobuyuki Omori), 岡村潤(Jun Okamura), 森辰則(Tatsunori Mori), 中川裕志(Hiroshi Nakagawa). t f ・ i d f 法を用いた関連マニュアル群のハイパーテキスト化. 情報処理学会研究報告FI-47-8/NL-121-16, 情報処理学会, Sep. 1997.

[3] Gerard Salton, Amit Singhal, Chris Buckley, and Mandar Mitra. Automatic text decomposition using text segments and text themes. In Proc. of Hypertext '96, pp. 53-65. the Association for Computing Machinery, Mar. 1996.

[4] Inderjeet Mani and Eric Bloedorn. Summarizing similarities and differences among related document. chapter 23, pp. 357-379. The MIT Press, London, 1999. (reprint of Information Processing and Management, Vol. 1, No. 1, pp. 1-23, 1999).

[5] 特開平 7 - 3 2 5 8 2 7

[6] 特開 2 0 0 0 - 5 7 1 5 2 (P 2 0 0 0 - 5 7 1 5 2 A)

[7] 特開平 1 1 - 3 9 3 3 4

このうち、文献 [1] では、関連論文に見られる一致点・相違点を、著者と命題(proposition)との2つの観点で一覧表の形にまとめた、“Synthesis Grid”というインタフェース(画面)が提案されている。

【 0 0 0 3 】

また、文書の関連箇所を抽出する従来の技術として、同一語彙の出現を手掛かりに、関連文書中の関連箇所にハイパーリンクを設定する技術が知られている。例えば、文献〔2〕では、文書中の一節に相当する「セグメント」を単位に文書を分割し、語彙的類似度の高いセグメント間にハイパーリンクを設定する技術が示されている。また、文献〔5〕や文献〔6〕では、関連文書中の同一キーワードの出現箇所にハイパーリンクを設定する技術等が示されている。

【0004】

その他、関連箇所の検出に関する技術として、文献〔3〕では、語彙的類似性の高い段落群を検出することで、単一文書中の関連箇所を抽出する技術が示されている。また、文献〔4〕では、文書中の語彙の連鎖等に基づく活性伝搬ネットワークを用いて、共通の関連語群を含む文等を検出する技術が示されている。

【0005】

また、関連文書の一致点・相違点を出力する技術としては、文献〔7〕に、複数の関連記事に対して、共通の内容とそれ以外の内容を区別して出力する技術が示されている。例えば、複数の入力記事のうち、1つを主記事としてその全体を表示し、残りの記事については、独自情報（共通情報以外の情報）のみを補足情報として表示する方法が示されており、また、主記事として全体を表示した記事に関しては、共通情報を強調表示（網かけして表示）する方法等も示されている。

【0006】

【発明が解決しようとする課題】

しかしながら、上述した従来の技術には、以下の2つの問題がある。

第1の問題は、関連箇所を認定する単位が固定的であるため、粒度の異なる話題に対して、適切な関連箇所を検出することが難しいということである。つまり、上述の従来技術では、節・段落・文（または語の出現位置そのもの）のいずれか1つに比較の単位を固定しているため、基本的に検出できるのは、節対節、段落対段落等、比較の単位の大きさの箇所同士に限られることになる。

【0007】

このため、例えば、第1の閲覧文書中で2段落からなる箇所が、ひとつのまと

まりとして、第 2 の閲覧文書中の数段落以上の大きさの箇所と関連している場合等には、関連箇所を対比できる形で適切に切り出すことが難しい。それを実現するためには、関連箇所として検出された箇所を併合する等、何らかの別の手段を講じることが必要になる。

【 0 0 0 8 】

第 2 の問題点は、ある話題に関する関連箇所と、別の話題に関する関連箇所あるいは元の文書全体との関係を十分に表現できないことである。例えば、長めで話題が複雑に組み入った文書同士を比較すると、複数の話題に関する関連箇所が、互いに絡み合っていて見いだされることがある。

【 0 0 0 9 】

このような場合、個々の話題についてそれぞれの文書から抽出した関連箇所を相互に比較するだけでなく、文書間に共通する複数の話題の相互関係や、各関連箇所が出現する文脈等も考慮して、子細に吟味する必要が生ずる。このとき、複数の関連箇所を一覧でき、各関連箇所の周辺部も容易に参照できることが望ましいが、上述の従来技術ではこのような機能は実現されていない。

【 0 0 1 0 】

本発明の第 1 の課題は、それぞれの文書毎に粒度の異なる話題に対しても適切な関連箇所を切り出して提示する文書閲覧装置およびその方法を提供することである。また、本発明の第 2 の課題は、複数の話題に関する関連箇所を、比較・分析しやすい形態で提示する文書閲覧装置およびその方法を提供することである。

【 0 0 1 1 】

【課題を解決するための手段】

図 1 は、本発明の文書閲覧装置の原理図である。図 1 の文書閲覧装置は、話題階層認定手段 1、関連話題抽出手段 2、および関連箇所提示手段 3 を備え、閲覧対象として指定された複数の文書を利用者に提示して、それらの文書の比較作業を支援する。

【 0 0 1 2 】

話題階層認定手段 1 は、複数の閲覧対象文書のそれぞれの話題階層を認定する。ここで、話題階層とは、文書を構成する複数の話題のまとまりが 2 段以上の階

層構造を成していることを意味する。この階層構造は、例えば、文書を構成する複数の大きな話題のまとまりの各々が、1つ以上のより小さな話題のまとまりを含み、小さな話題のまとまりの各々が、1つ以上のさらに小さな話題のまとまりを含むというような話題の包含関係に対応する。

【0013】

関連話題抽出手段2は、認定された話題階層に基づき、複数の閲覧対象文書に共通して現れる関連話題を抽出する。このとき、複数の文書にそれぞれ対応する複数の話題階層を比較し、関連性の強い話題の組み合わせを抽出し、複数の文書にまたがる関連話題として出力する。例えば、文書D1と文書D2から、第1の話題階層と第2の話題階層が得られた場合、第1の話題階層中の様々な粒度の話題それぞれを、第2の話題階層中の様々な粒度の話題それぞれと比較し、関連性の強い話題の対を抽出し、それらを関連話題として出力する。

【0014】

関連話題提示手段3は、抽出された関連話題に対応する記述箇所を、それぞれの閲覧対象文書から切り出して出力する。このとき、切り出された記述箇所は、複数の閲覧対象文書にまたがる関連箇所として出力される。

【0015】

このように、文書閲覧装置は、話題階層認定手段1によって、個々の閲覧対象文書に含まれる様々な粒度（大きさ）の話題を検出し、それに基づき、関連話題抽出手段2によって、文書間にまたがる関連話題を検出する。そして、関連話題提示手段3によって、関連話題に対応する文書に記述箇所を切り出して出力する。

【0016】

閲覧対象文書に含まれる様々な粒度の話題を単位に、網羅的に話題の関連性を調べることで、大きさの異なる記述箇所の対応関係を検出することができる。例えば、文書D1中の2段落相当の大きさの箇所が、ひとつのまとまりとして、文書D2中の数段落以上の大きさの箇所と関連している場合等にも、適切な関連箇所を切り出すことが可能になる。

【0017】

さらに、図 1 の文書閲覧装置は、以下のような様々な機能を有する。

関連話題抽出手段 2 は、関連話題の関連度を、各話題に対応する文書中の記述箇所の語彙的類似性によって求め、話題の包含関係に基づき設定した閾値により関連話題を選択する。例えば、上位層の話題 A と話題 B の対が関連度 R 1 を示している場合、話題 A と話題 B が包含するより小さな話題のいずれもが、いずれの話題とも R 1 以上の関連度を示していない場合にのみ、話題 A と話題 B の対を関連話題として出力する。

【 0 0 1 8 】

これにより、不適切な関連箇所の出力が抑制され、より効率的に関連箇所を出力できるようになる。例えば、2 つの閲覧対象文書のそれぞれに、複数段落からなる話題のまとまりがあり、それらが関連している場合、それらのまとまりを構成する一部の段落もまたある粒度の話題として並行的に関連していることがある。

【 0 0 1 9 】

すなわち、文書 D 1 中の第 1、第 2 段落と、文書 D 2 の第 1、第 2 段落の間で、2 段落のまとまり同士においても、また第 1 段落同士、第 2 段落同士においても、関連が見い出される場合等がある。このような場合にも、まとまり同士の関係として関連箇所を出力するか、個々の段落同士の関係として関連箇所を出力するかを、適切に選択し、冗長な出力を抑制することができる。

【 0 0 2 0 】

また、関連話題提示手段 3 は、関連箇所を関連話題単位に分け、対置して提示する。これにより、複数の関連話題が検出された場合においても、利用者は、個々の話題に関する対応箇所を対比して閲覧することが可能になる。

【 0 0 2 1 】

また、話題関連提示手段 3 は、各関連箇所の内容を短く要約して出力することもできる。これにより、多くの関連話題が検出された場合においても、利用者は、関連箇所全体の内容を一覧できるようになる。

【 0 0 2 2 】

また、関連話題提示手段 3 は、各関連箇所に対応する元の文書の部分を連動し

て提示することもできる。例えば、それぞれの関連箇所を元文書参照用のボタン（ハイパーリンク等）を添えて提示し、そのボタンによる要求に応じて、元文書の関連部を、併置した別のウィンドウに提示する。これにより、利用者は、各関連箇所が出現する文脈を参照しながら、関連箇所の内容を吟味すること等が可能になる。

【0023】

また、関連話題提示手段3は、閲覧対象文書の話題階層を表現した図面を合わせて提示し、図面上での利用者の指定に応じて、閲覧対象文書の対応箇所を連動して提示する。例えば、各話題をノードとする話題階層の木構造グラフを2つ併置し、また、関連話題をノード間のアークとして提示し、利用者の要求を受け付ける。そして、利用者がアークを指定した場合には、そのアークに対応する関連箇所を別のウィンドウに提示し、また、ノードを指定した場合には、そのノードに対応する部分を同様に提示する。

【0024】

これにより、利用者は、文書全体の話題の構成を手掛かりに、必要に応じて別の関連箇所の内容等を参照しながら、関連箇所を吟味することが可能になり、より効率的に複数文書を比較・閲覧できるようになる。

【0025】

また、関連話題提示手段3は、複数の閲覧対象文書の中の1つの文書を基準として、関連話題に関する記述箇所を他の閲覧対象文書から取り込むことで、新たな統合文書を作成し、提示する。これにより、例えば、利用者は、複数文書をまとめたレポート等の統合文書を効率的に作成できるようになる。

【0026】

【発明の実施の形態】

以下、図面を参照しながら、本発明の実施の形態を詳細に説明する。

本発明は、文書の一致点・相違点をわかりやすく提示するという機能に関し、現在の技術で自動化可能な範囲で実現したものである。具体的には、文書内の関連箇所の自動抽出技術を利用して、複数の関連文書における関連箇所を対比して提示する機能を実現している。

【 0 0 2 7 】

図 2 は、本発明の文書閲覧装置の基本構成を示している。図 2 の文書閲覧装置 1 2 は、入力部 (input unit) 2 1、単語認定部 (tokenizer) 2 2、単語辞書 (machine readable dictionary) 2 4、話題階層認定部 (thematic hierarchy detector) 2 5、関連話題抽出部 (related topic extractor) 2 7、および出力部 (output unit) 2 8 を備える。

【 0 0 2 8 】

図 1 の話題階層認定手段 1、関連話題抽出手段 2、および関連話題提示手段 3 は、それぞれ、図 2 の話題階層認定部 2 5、関連話題抽出部 2 7、および出力部 2 8 に対応する。

【 0 0 2 9 】

図 2 において、文書閲覧装置 1 2 は、複数の閲覧対象文書 1 1 が入力されると、それらの閲覧対象文書にまたがり、共通話題に関する関連箇所を抽出し、利用者 1 3 に提示する。

【 0 0 3 0 】

入力部 2 1 は、複数の閲覧対象文書 1 1 を読み込み、それぞれを逐次単語認定部 2 2 に渡す。単語認定部 2 2 は、サブモジュールの形態素解析部 (morphological analyzer) 2 3 を用いてそれぞれの文書 1 1 を言語的に解析して、文書 1 1 に含まれる内容語 (名詞・動詞・形容詞・形容動詞等) を切り出し、対応する文書 1 1 の部分に印を付ける。このとき、形態素解析部 2 3 は、単語辞書 2 4 を参照して、文書 1 1 中の文を、品詞情報付きの単語リストに変換する。単語辞書 2 4 は、形態素解析用の単語辞書であって、単語の表記文字列と品詞・活用の情報との対応関係等を記述している。

【 0 0 3 1 】

話題階層認定部 2 5 は、内容語の印付きの複数の閲覧対象文書 1 1 を受け取り、それぞれの文書 1 1 の話題階層を認定し、出力する。話題階層認定部 2 5 は、まず、サブモジュールの話題境界候補区間認定部 (thematic boundary detector) 2 6 を用いて、文書中の様々な粒度 (大きさ) の話題のまとまりを自動認定する。ここで、話題のまとまりとは、共通の話題について記述している文書の部分

を指す。そして、大きな話題のまとまりと小さな話題のまとまりを対応付けて話題階層データを作成し、出力する。

【 0 0 3 2 】

話題境界候補区間認定部 2 6 は、語彙的結束度の小さい区間を話題境界の候補区間として認定する。語彙的結束度とは、文書中の各位置の近傍領域における語彙的結束性の強さを表す指標であり、例えば、各位置の前後に設定したある幅の窓内に出現する語彙の類似性から求められる。

【 0 0 3 3 】

関連話題抽出部 2 7 は、複数の閲覧対象文書 1 1 のそれぞれに対応する、複数の話題階層を、話題階層認定部 2 5 から受け取り、2 つ以上の文書に共通して現れる関連話題を検出し、関連話題のリストを出力する。

【 0 0 3 4 】

出力部 2 8 は、関連話題抽出部 2 7 が抽出した関連話題に対応する記述箇所をそれぞれの文書から切り出し、関連話題毎にまとめて（または対応づけて）、利用者 1 3 に提示する。

【 0 0 3 5 】

図 2 の文書閲覧装置 1 2 は、例えば、図 3 に示すような情報処理装置（コンピュータ）を用いて構成することができる。図 3 の情報処理装置は、出力装置 4 1、入力装置 4 2、CPU（中央処理装置）4 3、ネットワーク接続装置 4 4、媒体駆動装置 4 5、補助記憶装置 4 6、およびメモリ（主記憶）4 7 を備え、それらはバス 4 8 により互いに接続されている。

【 0 0 3 6 】

メモリ 4 7 は、例えば、ROM（read only memory）、RAM（random access memory）等を含み、文書閲覧処理に用いられるプログラムとデータを格納する。ここでは、図 2 に示した入力部 2 1、単語認定部 2 2、形態素解析部 2 3、話題階層認定部 2 5、話題境界候補区間認定部 2 6、関連話題抽出部 2 7、および出力部 2 8 が、プログラムモジュールとして格納されている。CPU 4 3 は、メモリ 4 7 を利用してプログラムを実行することにより、必要な処理を行う。

【 0 0 3 7 】

出力装置 4 1 は、例えば、ディスプレイやプリンタ等であり、利用者 1 3 への問い合わせや、閲覧対象文書 1 1 および処理結果等の出力に用いられる。入力装置 4 2 は、例えば、キーボード、ポインティングデバイス、タッチパネル、スキャナー等であり、利用者 1 3 からの指示や閲覧対象文書 1 1 の入力に用いられる。

【 0 0 3 8 】

補助記憶装置 4 6 は、例えば、磁気ディスク装置、光ディスク装置、光磁気ディスク (magneto-optical disk) 装置等であり、閲覧対象文書 1 1、単語辞書 2 4 等の情報を格納する。また、情報処理装置は、この補助記憶装置 4 6 に、上述のプログラムとデータを保存しておき、必要に応じて、それらをメモリ 4 7 にロードして使用する。

【 0 0 3 9 】

媒体駆動装置 4 5 は、可搬記録媒体 4 9 を駆動し、その記録内容にアクセスする。可搬記録媒体 4 9 としては、メモ리카ード、フロッピーディスク、CD-R OM (compact disk read only memory)、光ディスク、光磁気ディスク等、任意のコンピュータ読み取り可能な記録媒体が用いられる。利用者 1 3 は、この可搬記録媒体 4 9 に上述のプログラムとデータを格納しておき、必要に応じて、それらをメモリ 4 7 にロードして使用する。

【 0 0 4 0 】

ネットワーク接続装置 4 4 は、LAN (local area network) 等の任意のネットワーク (回線) を介して外部の装置と通信し、通信に伴うデータ変換を行う。また、情報処理装置は、上述のプログラムとデータをネットワーク接続装置 4 4 を介して、サーバ等の他の装置から受け取り、必要に応じて、それらをメモリ 4 7 にロードして使用する。

【 0 0 4 1 】

図 4 は、図 3 の情報処理装置にプログラムとデータを供給することのできるコンピュータ読み取り可能な記録媒体を示している。可搬記録媒体 4 9 やサーバ 5 0 のデータベース 5 1 に保存されたプログラムとデータは、メモリ 4 7 にロードされる。そして、CPU 4 3 は、そのデータを用いてそのプログラムを実行し、

必要な処理を行う。このとき、サーバ 5 0 は、プログラムとデータを伝送する伝搬信号を生成し、ネットワーク上の任意の伝送媒体を介して、情報処理装置に送信する。

【 0 0 4 2 】

次に、図 2 に示した文書閲覧装置 1 2 の各モジュールの動作を、具体例を用いてより詳細に説明する。

閲覧対象文書の例としては、「第 1 4 9 回衆議院本会議会議録第 2 号」（2 0 0 0 年 7 月 3 1 日）から、水島広子議員による代表質問（第 1 の閲覧対象文書）とそれに対する首相の答弁（第 2 の閲覧対象文書）を、それぞれ 1 つの文書として切り出したものを用いた。衆議院の代表質問は、党を代表する議員がいくつかの項目を一括して質問した後、首相・関係大臣が答弁する形で進められるが、この代表質問では、子供の教育、民法改正、国会運営、有害情報、小児医療、歳費支給方式の 6 つの問題に関し、計 8 項目が質問されている。

【 0 0 4 3 】

図 5 は、代表質問部を切り出した第 1 の閲覧対象文書の冒頭部である。なお、図 5 において、下線を引いた箇所、すなわち、冒頭の発言議員名および括弧で囲まれた議事進行に関する補足情報については、代表質問の内容ではないため、除外してから、以降の処理を行った。首相の答弁部を切り出した第 2 の閲覧対象文書においても、同様に冒頭の発言者名と括弧内の補足情報は除去して処理を行っている。

【 0 0 4 4 】

図 6 は、単語認定部 2 2 による単語認定処理のフローチャートである。単語認定部 2 2 は、まず、個々の閲覧対象文書に形態素解析を施し、品詞付きの単語リストを作成する（ステップ S 1 1）。次に、品詞を手掛かりに内容語（名詞・動詞・形容詞・形容動詞）を認定し、内容語に対応する文書の部分に印を付けて（ステップ S 1 2）、処理を終了する。図 7 は、図 5 の文書部分に対する単語認定部 2 2 の処理結果を示している。

【 0 0 4 5 】

図 6 のステップ S 1 1 において、形態素解析部 2 3 は、図 8 に示すような形態

素解析処理を行う。形態素解析部 2 3 は、まず、単語リストをクリアし（ステップ S 2 1）、文書の先頭から句点（またはピリオド）等を手掛かりに文の取り出しを試み（ステップ S 2 2）、文が取り出せたかどうかを判定する（ステップ S 2 3）。

【 0 0 4 6 】

文が取り出せれば、次に、単語辞書 2 4 を参照して、文に含まれている単語の候補を求める（ステップ S 2 4）。日本語の場合は、図 7 に示したように、単語と単語の境界が形式的に明示されていないので、文に含まれる部分文字列に対応するすべての単語を候補として求める。例えば、「東京都は大都市だ」という文が取り出された場合、図 9 に示すように、この文に含まれるすべての部分文字列が単語の候補となる。

【 0 0 4 7 】

これに対して、英語の場合は、単語の境界が空白（スペース）により明示されているため、空白で区切られた文字列に対応する単語について、品詞の候補を求めることが主な処理となる。例えば、“Tokyo is the Japanese capital.” という文が取り出された場合、図 1 0 に示すように、この文に明示的に含まれる 5 つの単語の基本形と品詞が求められる。

【 0 0 4 8 】

次に、形態素解析部 2 3 は、品詞レベルの接続の観点から、妥当な単語の並びを選択し（ステップ S 2 5）、選択された単語の並びに品詞と出現位置の情報を付加して、出現順に単語リストに追加する（ステップ S 2 6）。次に、次の文の取り出しを試み（ステップ S 2 7）、ステップ S 2 3 以降の処理を繰り返す。そして、ステップ S 2 3 において文が取り出せなくなると、処理を終了する。

【 0 0 4 9 】

図 1 0 の単語認定結果において、墨付き括弧で括られた部分が形態素解析部 2 3 の認定した内容語である。内容語が活用語（動詞・形容詞）の場合、墨付き括弧内で、スラッシュ（／）の前の部分は語幹を表し、スラッシュの後の部分は終止形の活用語尾を表す。これは、後の処理で単語の区別を行うために用いられる情報であるが、この情報の代わりに、品詞と活用を付加しておいてもよい。要す

るに、例えば、「い／う」と「い／る」のように、語幹だけでは区別の付かない単語を区別するための識別情報であれば、任意のものを用いることができる。

【0050】

また、ステップS25において、単語の並びの妥当性を評価する方法は、形態素解析法として各種のものが知られており、任意のものを用いることができる。例えば、以下の文献[8]、[9]、および[10]では、単語の並びの妥当性を訓練データにより推定された出現確率を用いて評価する方法が報告されている。

[8] Eugene Charniak. Hidden markov models and two applications. In Statistical Language Learning, chapter 3, pp. 37-73. The MIT Press, 1993.

[9] 永田昌明. 前向きDP後向きA^{*} アルゴリズムを用いた確率的日本語形態素解析システム. 情報処理学会研究報告NL-101-10, 情報処理学会, May 1994.

[10] Masaaki Nagata. A stochastic japanese morphological analyzer using a forward-DP backward-A^{*} N-best search algorithm. In Proc. of COLING '94, pp. 201-207, Aug. 1994.

なお、図7の例では、単語認定部22がすべての内容語を切り出しているが、切り出しの対象を名詞だけに絞っても構わない。また、英語の文書を対象に処理する場合には、形態素解析処理を行う代わりに、空白で区切られたすべての語のうち、話題に関わらずどこにでも出現する語彙（冠詞、前置詞等の機能語や特に高い頻度で出現する語）を取り除いて、単語を切り出してもよい。このような処理は、単語辞書24の代わりに、機能語や特に高い頻度で出現する語を格納したストップワードリスト（stop word list）を用意すれば、容易に実現できる。

【0051】

次に、話題階層認定部25の処理について説明する。本実施形態においては、話題のまとまりは、先願の特開平11-272699「文書要約装置およびその方法」に示された技術に基づいて認定している。この方法では、以下の手順で話題の階層構成が認定される。

1. 話題境界位置の区間推定

ある窓幅で計算した結束度に基づき、話題境界が存在しそうな位置を、話題境

界候補区間として求める。そして、大きさの異なる複数の窓幅に対してこの処理を繰り返し、大きな話題の切れ目を示す境界から小さな話題の切れ目を示す境界まで、話題の大きさ別に話題境界候補区間を求める。

2. 話題の階層関係の認定

異なる窓幅により求めた話題境界候補区間を統合し、話題の階層構造とそれぞれの話題境界の位置を決定する。

【0052】

図11は、話題階層認定部25による話題階層認定処理のフローチャートである。話題階層認定部25は、まず、最大窓幅 w_1 、最小窓幅 w_{min} 、窓幅比 r の3つのパラメータを利用者から受け取り（ステップS41）、結束度を測定するための窓幅の集合 W を求める（ステップS42）。窓幅の集合 W は、初項を w_1 とし、公比を $1/r$ とする等比級数から、 w_{min} 以上の大きさの項を集めて作成される。

【0053】

このとき、 W における最大窓幅 w_1 としては、文書全体のサイズの $1/2 \sim 1/4$ 程度の大きさを与え、最小窓幅 w_{min} としては、段落程度の大きさ（例えば40語）を与え、窓幅比 r としては2を与えておけば、実用上十分である。以下では、 $w_1 = 320$ （語）、 $w_{min} = 40$ （語）、 $r = 2$ の値を用いている。

【0054】

次に、話題階層認定部25は、図7に示したように、内容語に印が付けられた文書をもとに、文書中の各位置の結束度を、 W 中のそれぞれの窓幅毎に計算し、結束度系列として記録する（ステップS43）。

【0055】

ここでは、まず、文書の各位置（基準点）の前後に設定した2つの窓の中に出現している語彙（ここでは内容語）を比較し、共通している語彙が多い程大きくなるような値を計算して、その位置における結束度とする。そして、窓の位置を文書の冒頭から末尾に向かって一定の刻み幅 t_{ic} でずらしながら、結束度の計算を繰り返し、計算した結束度を、文書の冒頭から末尾に向かう系列として記録

する。

【0056】

刻み幅 t_{ic} は、窓幅より小さければいずれの値でも構わないが、処理効率を考慮して、ここでは、窓幅の $1/8$ の値を用いた。この t_{ic} の値は、利用者により指定することも可能である。

【0057】

結束度の計算方法としては各種の方法が考えられるが、以下では、情報検索等の分野で類似度の尺度として広く用いられてきた余弦測度 (cosine measure) を用いている。余弦測度は次の式により求められる。

【0058】

【数1】

$$\text{sim}(b_l, b_r) = \frac{\sum_t w_{t,b_l} w_{t,b_r}}{\sqrt{\sum_t w_{t,b_l}^2 \sum_t w_{t,b_r}^2}}$$

【0059】

ここで、 b_l と b_r は、それぞれ、左窓（文書の冒頭側の窓）、右窓（文書の末尾側の窓）に含まれる文書の部分を表し、 w_{t,b_l} 、 w_{t,b_r} は、それぞれ、左窓、右窓に出現する単語 t の出現頻度を表す。また、(1) 式の右辺の \sum_t は、単語 t に関する総和を表す。

【0060】

(1) 式の類似度は、左右の窓に含まれる語彙に共通の人が多いほど大きくなり（最大1）、共通のものが無い時に0となる。つまり、この値が大きい部分は、左右の窓で共通の話題を扱っている可能性が高く、逆に、この値が小さい部分は、話題の境界である可能性が高いことになる。

【0061】

次に、図12は、ステップS43で記録された結束度の系列の例を示している。図12では、簡単のため、窓幅 w の $1/4$ を刻み幅 t_{ic} として用いており、文書領域 $a_1 \sim a_{11}$ は、刻み幅 t_{ic} に対応する一定幅の領域である。また、

c 1 は、文書中の a 4 と a 5 の境界を基準点として計算した、窓幅 w の結束度を表す。すなわち、c 1 は、文書領域 a 1 ～ a 4 の部分を左窓の範囲とし、a 5 ～ a 8 の部分を右窓の範囲として計算された結束度である。

【0062】

次の c 2 は、窓を t i c 分だけ右へずらして計算された結束度を表し、a 5 と a 6 の境界を基準点とする窓幅 w の結束度である。このようにして、窓を t i c 分ずつ順に右へずらして計算した c 1, c 2, c 3, c 4, . . . を、文書の冒頭から末尾へ向かう窓幅 w の結束度系列と呼ぶ。

【0063】

図 1 3 は、上述の単語認定結果において、文書の冒頭から各基準点までの間に出現した内容語の延べ数を横軸にとり、最小窓幅（40 語）の結束度系列をプロットしたグラフである。例えば、図 1 2 の結束度 c 2 の場合は、a 1 ～ a 5 の領域中の内容語の延べ数が、文書における基準点の位置となる。ここでは、40 語の窓幅の $1/8$ （5 語）を刻み幅 t i c として、文書の冒頭から末尾に向かって結束度を計算している。

【0064】

次に、話題階層認定部 2 5 は、サブモジュールの話題境界候補区間認定部 2 6 を使っぶし、大きな話題の切れ目を示す境界から小さな話題の切れ目を示す境界まで、話題の大きさ別に話題境界候補区間を求める。

2. 話題の階層関係の認定

異なる窓幅により求めた話題境界候補区間を統合し、話題の階層構造とそれぞれの話題境界の位置を決定する。

【0065】

図 1 1 は、話題階層認定部 2 5 による話題階層認定処理のフローチャートである。話題階層認定部 2 5 は、まず、最大窓幅 w 1、最小窓幅 w__m i n、窓幅比 r の 3 つのパラメータを利用者から受け取り（ステップ S 4 1）、結束度を測定するための窓幅の集合 W を求める（ステップ S 4 2）。窓幅の集合 W は、初項を w 1 とし、公比を $1/r$ とする等比級数から、w__m i n 以上の大きさの項を集めて作成される。

【0066】

このとき、Wにおける最大窓幅 w_1 としては、文書全体のサイズの $1/2 \sim 1/4$ 程度の大きさを与え、最小窓幅 w_{min} としては、段落程度の大きさ（例えば40語）を与え、窓幅比 r としては2を与えておけば、実用上十分である。題の開始位置）を求めることも可能である。

【0067】

次に、図11のステップS44における話題境界候補区間認定処理について、図12および図14を使って説明する。ここで用いられる移動平均法は、株価の変動などの統計的分析方法である時系列分析（time series analysis）において、細かい変動を取り除いて大局的な傾向を把握するために使われている。本実施形態では、結束度系列の移動平均値を細かい変動を無視するために用いるだけでなく、それを移動平均の開始点における順方向結束力および移動平均の終了点における逆方向結束力とみなすことで、話題境界候補区間認定のための直接的な手掛かりとしている。

【0068】

図12は、前述したように、結束度の系列 $c_1 \sim c_4$ と文書領域 $a_1 \sim a_{11}$ との関係を示している。結束度系列の移動平均値とは、例えば、 $(c_1 + c_2) / 2$ （2項の移動平均）、 $(c_1 + c_2 + c_3) / 3$ （3項の移動平均）、 $(c_1 + c_2 + c_3 + c_4) / 4$ （4項の移動平均）のように、結束度系列において連続する n 個の値を算術平均した値である。

【0069】

図14は、図12の結束度系列の移動平均の例と文書領域との関係を示している。ここでは、移動平均の例として、図12の結束度の2項～4項の移動平均が示され、それぞれの移動平均に関わる結束度の計算において、各文書領域が使用された回数が示されている。このうち、下線を付けた値は、対応する文書領域が移動平均に関わるすべての結束度の計算に用いられていることを表す。

【0070】

例えば、左上角の値“1”は、 $c_1 \sim c_4$ までの4項の移動平均において、文書領域 a_1 が一度だけ左窓の一部として扱われたことを示している。また、その

右の値“2”は、c1～c4までの4項の移動平均において、文書領域a2が2回左窓の一部として扱われたことを示している。他の使用回数についても、同様である。

【0071】

結束度は境界の前後の部分の結び付きの強さを表す指標であるので、領域a1を左窓に含んで得られた結束度c1を用いて計算された移動平均値も、領域a1がそれより右側（文書の末尾側）の部分に結び付いているかどうかを示す指標の1つと考えられる。

【0072】

言い換えれば、移動平均値は、移動平均をとった結束度の左窓部分の領域（c1～c4の4項平均に対してはa1～a7）が文書の末尾へ向かう方向（順方向：図15では右方向）に引っ張られる強さの指標（順方向結束力）になっていると言える。一方、逆に、移動平均をとった結束度の右窓部分の領域（c1～c4の4項平均に対してa5～a11）が文章の冒頭方向（逆方向：図15では左方向）に引っ張られる強さの指標（逆方向結束力）になっているとも言える。

【0073】

ここで、結束力とそれぞれの文書領域との関連性を考察すると、結束度の計算においてより多く窓に含まれていた領域との関連が強いと考えられる。また、語彙的結束性は、一般に、近傍で繰り返される語彙に基づくものほど強いと考えられるので、移動平均をとった結束度の基準点（左右の窓の境界位置）に近い位置にある領域ほど関連が強いとも言える。

【0074】

例えば、図14の4項の移動平均については、結束度の基準点は、a4とa5の境界、a5とa6の境界、a6とa7の境界、およびa7とa8の境界の4つである。この場合、a4は最も多く左窓に含まれており、かつ、これらの基準点に最も近いことが分かる。また、a8は最も多く右窓に含まれており、かつ、これらの基準点に最も近いことが分かる。したがって、移動平均値と最も関連の強い領域は、左窓についてはa4、右窓についてはa8となる。

【0075】

同様にして、3項の移動平均と最も関連の強い領域を選ぶと、左窓についてはa4、右窓についてはa7となり、2項の移動平均と最も関連の強い領域を選ぶと、左窓についてはa4、右窓についてはa6となる。これらの領域の使用回数は、図14では太線の枠で囲んで示されている。

【0076】

以上の考察に基づき、話題境界候補区間認定部26は、結束度の移動平均値を、移動平均をとった領域内の最初の基準点における順方向結束力および最後の基準点における逆方向結束力の指標として取り扱う。例えば、c1～c4の4項の移動平均値は、a4とa5の境界における順方向結束力およびa7とa8の境界における逆方向結束力となる。

【0077】

図15は、話題境界候補区間認定部26による話題境界候補区間認定処理のフローチャートである。候補区間認定部26は、まず、話題階層認定部25から結束度系列の刻み幅ticを受け取り、利用者から移動平均の項数nを受け取る（ステップS51）。

【0078】

これらのパラメータの値の目安は、刻み幅ticについては、例えば、窓幅wの $1/8 \sim 1/10$ 程度の大きさであり、項数nについては、 w/tic の半分（4～5）程度である。また、移動平均をとる領域の最初の基準点から最後の基準点までの隔たりを、 $(n-1) * tic$ により計算して、それを移動平均の幅d（語）とする。

【0079】

次に、文書中の各位置pについて、 $p \sim p+d$ の範囲内で結束度の移動平均をとり、平均値を位置pにおける順方向結束力として記録する（ステップS52）。この値は、同時に、移動平均をとった範囲の終了位置 $p+d$ における逆方向結束力としても記録される。

【0080】

次に、記録された順方向結束力をもとに、文書中の冒頭から末尾に向かって各位置における順方向結束力と逆方向結束力の差（順方向結束力－逆方向結束力）

を調べ、その値が負から正に変化する位置を負の結束力拮抗点 m_p として記録する（ステップ S 5 3）。

【0081】

負の結束力拮抗点とは、その位置の左では逆方向結束力が優勢であり、その位置の右では順方向結束力が優勢であるような点である。したがって、この点の左右の部分は意味的な結び付きが弱いと考えられ、負の結束力拮抗点は話題境界の候補位置となる。

【0082】

次に、記録された負の結束力拮抗点 m_p の直前・直後の d 語以内の範囲 $[m_p - d, m_p + d]$ を話題境界候補区間と認定して（ステップ S 5 3）、処理を終了する。

【0083】

ここで、順・逆方向の結束力の差に基づいて話題境界候補区間を認定する意味を、図 1 6 を使って説明する。図 1 6 は、図 1 3 の 4 0 0 語付近（3 7 0 語～4 0 0 語）における 4 0 語幅の窓による結束度と順・逆方向の結束力の分布を示している。刻み幅 t_{ic} としては、窓幅の $1/8$ を採用している。

【0084】

図 1 6 において、記号 + でプロットした折れ線グラフは、結束度 C の系列を表し、記号 * でプロットした折れ線グラフは、順方向結束力 FC の系列を表し、記号 \square でプロットした折れ線グラフは、逆方向結束力 BC の系列を表す。話題境界候補区間を表す矩形で示された領域については、後述することにする。

【0085】

また、点線で示された e_{p1} 、 e_{p2} 、 e_{p3} は、順・逆方向の結束力の差が 0 になる 3 つの点（結束力拮抗点）を表す。最初の点 e_{p1} の左側では、逆方向結束力が順方向結束力より優勢であり、その右側から次の点 e_{p2} までは、順方向結束力が逆方向結束力より優勢である。さらに、その右側から最後の点 e_{p3} までは、逆方向結束力が順方向結束力より優勢であり、その右側では、順方向結束力が逆方向結束力より優勢である。

【0086】

したがって、 $e p 1$ と $e p 3$ は、順方向結束力と逆方向結束力の差が負から正に変化する負の結束力拮抗点であり、 $e p 2$ は、その差が正から負に変化する正の結束力拮抗点である。

【 0 0 8 7 】

このような結束力の変化から、最初の点 $e p 1$ の左側の領域は、それより左側のいずれかの部分と比較的強い結束性を示しており、真中の点 $e p 2$ の両側の領域は、 $e p 2$ に向かって強い結束性を示しており、最後の点 $e p 3$ の右側の領域は、それより右側のいずれかの部分と比較的強い結束性を示していることが分かる。実際、順・逆方向の結束力と共にプロットした結束度は、 $e p 1$ と $e p 3$ の近傍で極小値をとり、 $e p 2$ の近傍で極大値をとっている。このように、順・逆方向の結束力の変化と結束度の変化は密接に関連している。

【 0 0 8 8 】

例えば、図 1 6 の結束力拮抗点 $e p 3$ の近傍には、結束度の極小点（この場合は $c 3$ ）がある。上向き矢印で示した FC 、 BC の極小値は、水平矢印の部分の結束度（ $c 1 \sim c 4$ の4項）を移動平均した値である。このように、通常は、結束度の極小点に対応して、その近傍（移動平均の幅以内）で、結束力も極小値をとる。ただし、移動平均をとる領域より狭い範囲で細かい変動がある場合には、移動平均の平滑化作用により、移動平均値すなわち結束力が極小値をとらないこともある。

【 0 0 8 9 】

また、順方向結束力は移動平均値を移動平均をとる領域の開始位置に記録した指標であるので、順方向結束力の極小位置は結束度の極小位置の左になる。同様の理由により、逆方向結束力の極小位置は結束度の極小位置の右になる。そして、結束度の変動が十分に大きければ、移動平均をとる領域内に結束力拮抗点が生じることになる。

【 0 0 9 0 】

図 1 7 は、図 1 1 のステップ S 4 5 において行われる話題境界認定処理のフローチャートである。話題階層認定部 2 5 は、まず、認定された話題境界候補区間を、認定に使った結束度系列の窓幅と、話題境界候補区間内の結束力拮抗点の文

書における出現位置とによってソートしてまとめ、話題境界候補区間データの系列 $B(j)[p]$ を作成する (ステップ S61)。

【0091】

ここで、制御変数 j は、窓幅 w_j の結束度系列により認定されたことを表す系列番号であり、制御変数 p は、系列内の各話題境界候補区間を表すデータ番号である。実際には、 j は、窓幅の大きい順に 1, 2, ... の値をとり、 p は、結束力拮抗点の出現順に 1, 2, ... の値をとる。それぞれのデータ $B(j)[p]$ は、次のような要素データを含む。

【0092】

・ $B(j)[p].range$: 話題境界候補区間。(開始位置、終了位置)の組。

・ $B(j)[p].ep$: 結束力拮抗点。

【0093】

・ $B(j)[p].child$: 境界位置の話題境界候補区間の範囲で一致する $B(j+1)$ 系列の話題境界候補区間 (子候補区間)。

なお、結束力拮抗点は理論的には点であるが、前述のように、順方向結束力と逆方向結束力の差の符号が反転する地点を拮抗点として認定しているので、実際には、差が負の点 (開始位置) と差が正の点 (終了位置) の組により表される。そこで、本実施形態では、結束力拮抗点の開始位置 l_p と終了位置 r_p における (順方向結束力 - 逆方向結束力) の値を、それぞれ、 $DC(l_p)$ と $DC(r_p)$ として、左右の結束力が 0 になる点 e_p を、次式により補間して求める。

$$e_p = \frac{(DC(r_p) * l_p - DC(l_p) * r_p)}{(DC(r_p) - DC(l_p))} \quad (2)$$

そして、得られた e_p を、 $B(j)[p].ep$ とする。

【0094】

次に、話題階層認定部 25 は、窓幅の異なる話題境界候補区間データを関連付ける処理を行う。ここでは、1 つの系列に属する $B(j)[p]$ をまとめて B (

j) と記し、さらに、次のような表記法を用いて、以下の処理を説明する。

【0095】

- ・ i_e : 最小窓幅 w_{min} に対応する系列番号。
- ・ $|B(j)|$: $B(j)$ におけるデータ番号 p の最大値。

まず、処理対象を表す系列番号 i を 1 に初期化する (ステップ S62)。これにより、最大窓幅 w_1 による話題境界候補区間の系列が処理対象に設定される。そして、 $j+1 \leq j_e$ である限り、 j をインクリメントしながら、 $B(j+1)$ を関連付け対象の系列とする関連付け処理を行う。

【0096】

この関連付け処理では、処理対象系列中のそれぞれの話題境界候補区間データ $B(j)[p]$ ($p=1, \dots, |B(j)|$) について、それと同じ付近を境界候補としている関連付け対象系列中のデータ $B(j+1)[q]$ のうち、 $B(j+1)[q].e_p$ が $B(j)[p].e_p$ に最も近いデータが選ばれ、関連境界候補区間データとして $B(j)[p].child$ に格納される。

【0097】

具体的な手順は以下の通りである。まず、 $j+1$ と j_e を比較し (ステップ S63)、 $j+1 \leq j_e$ であれば、 p に 1 を代入して (ステップ S64)、 p と $|B(j)|$ を比較する (ステップ S65)。 $p \leq |B(j)|$ であれば、ステップ S66 以降の関連付け処理を行い、 p が $|B(j)|$ を越えれば、 $j = j+1$ において (ステップ S71)、ステップ S63 以降の処理を繰り返す。

【0098】

ステップ S66 では、話題階層認定部 25 は、関連付け候補となる系列中のデータ $B(j+1)[q]$ ($q=1, \dots, |B(j+1)|$) の中から、 $B(j+1)[q].e_p \in B(j)[p].range$ となるデータで $B(j+1)[q].e_p$ が $B(j)[p].e_p$ に最も近いデータを、関連付け対象データとして選択し、 $B(j)[p].child$ に格納する。

【0099】

ここで、 $B(j+1)[q].e_p \in B(j)[p].range$ という条件は、 $B(j)[p]$ の話題境界候補区間の中に $B(j+1)[q]$ の結束力拮抗

点が含まれていることを表す。

【0100】

図18は、関連付け対象データの選択例を示している。図18において、記号+でプロットした折れ線グラフは、処理対象に対応する80語幅の窓による順方向結束力の系列を表し、記号×でプロットした折れ線グラフは、80語幅の窓による逆方向結束力の系列を表す。また、記号*でプロットした折れ線グラフは、関連付け対象に対応する40語幅の窓による順方向結束力の系列を表し、記号□でプロットした折れ線グラフは、40語幅の窓による逆方向結束力の系列を表す。また、矩形で示された領域は、話題境界候補区間に対応し、点線で示された e_{p1} 、 e_{p3} は、40語幅の窓による結束力の拮抗点に対応する。

【0101】

例えば、処理対象データを $B(3)[4]$ とすると、その近傍には結束力拮抗点 e_{p1} 、 e_{p3} があり、それらに対応して関連付け対象系列の2つのデータ $B(4)[6]$ と $B(4)[7]$ がある。このうち、 $B(4)[7]$ の結束力拮抗点 e_{p3} は、 $B(3)[4]$ の話題境界候補区間（上方の矩形）に含まれているので、 $B(4)[7]$ が関連付け対象データとして選択される。

【0102】

次に、話題階層認定部25は、関連付け対象データが選択できたかどうかを判定し（ステップS67）、関連付け対象データが選択できた場合は、 $p = p + 1$ とおいて（ステップS70）、ステップS65以降の処理を繰り返す。

【0103】

条件を満たす関連付け対象データが見つからなかった場合には、 $B(j)[p]$ と同じ話題境界候補区間を持つ疑似的な関連付け対象データ（ダミーデータ） $B(j+1)[q]$ を作成し、 $B(j+1)$ の系列に挿入する（ステップS68）。

【0104】

ステップS68では、まず、 $B(j+1)[q].range$ と $B(j+1)[q].ep$ に、それぞれ、 $B(j)[p].range$ と $B(j)[p].ep$ の値を設定して、新たなデータ $B(j+1)[q]$ を作成する。そして、系列

$B(j+1)$ の中で、 $B(j+1)[q-1].ep < B(j+1)[q].ep$ かつ $B(j+1)[q].ep < B(j+1)[q+1].ep$ となるような位置に、作成したデータ $B(j+1)[q]$ を挿入する。

【0105】

これにより、疑似的な関連付け対象データのデータ番号 q が決定され、それ以降の既存データのデータ番号は書き換えられる。ここで、疑似的な話題境界候補区間データを作成するのは、以降の処理において、 j 以下のすべての系列番号の話題階層において話題文を抽出するためである。

【0106】

次に、作成されたダミーデータ $B(j+1)[q]$ を $B(j)[p].child$ に格納し（ステップS69）、ステップS70以降の処理を行う。そして、ステップS63において $j+1$ が je を越えれば、処理を終了する。

【0107】

最終的には、 je 未満のすべての系列番号 j のデータに対して、その話題境界候補区間内に結束力拮抗点を持つ系列番号 $j+1$ のデータが、 $B(j)[p].child$ に設定される。したがって、 $B(j)[p].child$ により、複数の階層の話題境界候補区間データが連鎖的に関連付けられることになる。

【0108】

図19は、こうして得られた話題境界の認定結果を示している。図19において、320語、160語、80語、40語の各窓幅（縦軸）の高さまで伸びた棒グラフが、それぞれの窓幅に対応する粒度の話題の最終的な話題境界、すなわち最小窓幅（40語）の結束力拮抗点の位置を表している。棒グラフと交差している矩形領域は、各窓幅の結束力により認定された話題境界候補区間を表す。

【0109】

図11のステップS46では、図19に示された話題境界を微調整して文の開始位置に合わせてから、各境界の間を1つの話題とする話題階層を作成する。この微調整により図19の話題境界の一部がシフトし、その結果、図20に示すような木構造の話題階層が生成される。

【0110】

例えば、図19の最小窓幅に対応する境界からは、矢印で示された15個の区画に対応して、図20の15個の話題が最下層の話題として認定される。また、80語の窓幅に対応する境界からは、15個の話題のうち、区画2と区画3、区画4から区画6まで、区画11と区画12、区画13と区画14の4つのグループに対応する話題をグループ毎に統合して得られる、計10個の話題が第2層の話題として認定される。

【0111】

図20の話題階層において、矩形で表されたノードは、認定されたそれぞれの話題に対応し、矩形内の数字は、図19の区画の番号に対応する。第2の閲覧対象文書に対しても同様の処理を行うことで、図21のような話題階層が生成される。

【0112】

次に、関連話題抽出部27の処理について説明する。図22は、関連話題抽出部27による関連話題抽出処理のフローチャートである。話題文抽出部27は、まず、第1の閲覧対象文書の話題階層T1と第2の閲覧対象文書の話題階層T2の2つの話題階層を入力として受け取る（ステップS101）。そして、話題階層T1中の任意の話題t1と話題階層T2中の任意の話題t2からなる話題対（t1, t2）のすべてについて、関連度を計算する（ステップS102）。

【0113】

本実施形態では、話題t1と話題t2の間の関連度R(t1, t2)を、t1、t2のそれぞれに対応する文書の区画s1、s2に含まれる語彙の類似性により求める。具体的には、R(t1, t2)は次式により計算される。

【0114】

【数2】

$$R(t1, t2) \equiv R(s1, s2) = \frac{\sum_t w_{t, s1} w_{t, s2}}{\sqrt{\sum_t w_{t, s1}^2 \sum_t w_{t, s2}^2}}$$

【0115】

ここで、 $w_{t,s1}$ 、 $w_{t,s2}$ は、それぞれ、区画 $s1$ 、 $s2$ における単語 t の重要度に相当する重みを表し、次式により計算される。

【0116】

【数3】

$$w_{t,s} = tf_{t,s} \times \log\left(\frac{|D|}{df_t}\right)$$

【0117】

(4) 式において、 $tf_{t,s}$ は、区画 s における単語 t の出現頻度を表し、 $|D|$ は、区画 s を含む文書を固定幅 (80 語) 刻みに区切って得られるブロックの数を表し、 df_t は、単語 t が出現しているブロックの数を表す。

【0118】

(3)、(4) 式は、情報検索分野で検索対象文書と質問文との関連度計算等でよく使われる、 $tf \times idf$ 法と呼ばれる計算法の変形例である。 $tf \times idf$ 法では、(4) 式の $|D| / df_t$ の部分を、文書内の区画ではなく、検索対象文書集合に含まれる文書を単位に計算する。すなわち、 $|D|$ を検索対象文書集合中の文書数とし、 df_t を単語 t が出現する文書数とすると、これらの式は通常の $tf \times idf$ 法の計算式となる。

【0119】

関連度 $R(t1, t2)$ は $tf \times idf$ 法により求めてもよいが、本実施形態の (3)、(4) 式によれば、閲覧対象文書だけから関連度を計算することができ、また、後述するように、これらの計算式で十分に有効な結果が得られているため、ここではこの計算法を採用している。

【0120】

関連話題抽出部 27 は、次に、第 1 の閲覧対象文書中の話題 $t1$ と第 2 の閲覧対象文書中の話題 $t2$ のすべてに対して、話題階層を利用しながら、話題対選別の閾値を求める。閾値としては、例えば、話題階層の部分木中の最大関連度が用いられる。ここで、ある話題 t に対する話題階層の部分木中の最大関連度とは

、 t もしくは話題階層における t の子孫 (t を構成するいずれかのより小さい話題) に対して計算された関連度の最大値のことである。

【0121】

関連話題抽出部 27 は、まず、話題 t_1 について最大関連度を求め、それを $t_1. \max$ に記録し (ステップ S103)、次に、話題 t_2 についても同様に、最大関連度を $t_2. \max$ に記録する (ステップ S104)。そして、 $T \equiv \{ (t_1, t_2) \mid R(t_1, t_2) \geq \max(t_1. \max, t_2. \max) \}$ により定義される話題対の集合 T を求めて、関連話題として出力し (ステップ S105)、処理を終了する。

【0122】

ここで、最大関連度に基づく関連話題抽出処理の具体例を、図 23 および図 24 を使って説明する。図 23 は、図 22 のステップ S102 における関連度の計算結果を示している。ここでは、関連度が 0.25 以上の値を持つ話題対が点線のアークで示されており、アークに添えられた数値が関連度を表す。また、2 つの木構造グラフのうち、左のグラフが図 20 の話題階層に対応し、右のグラフが図 21 の話題階層に対応する。

【0123】

ここで、右のグラフの右下角のノード (以下では、「ノード右 7」のように表記する) に着目する。このノードは、第 2 の閲覧対象文書の最後の最小区画に対応する話題を表すノードであり、グラフ上では、末端ノード (子ノードを持たないノード) となっている。

【0124】

したがって、このノードにおける最大関連度は、このノードに直接結びつけられたアークの関連度の最大値である。ノード右 7 では、(ノード左 13-14, ノード右 7) の話題対の関連度 0.35 が最大関連度となる。そして、ノード左 13-14 からは、0.35 を超える関連度をもつアークはないので、(ノード左 13-14, ノード右 7) の話題対は関連話題として出力される。

【0125】

一方、ノード右 6-7 に着目すると、このノード以下の部分木にノード右 7 が

含まれているので、ノード右6-7に直結しているアークの話題対は、少なくともノード右7の最大関連度（0.35）以上でなければ、関連話題として出力されない。ノード右6-7はこのようなアークを持たないので、ノード右6-7を含む話題対は関連話題としては出力されないことになる。

【0126】

以上のように、部分木における最大関連度を基準に話題対を選別することで、2つの閲覧対象文書間の関連話題を、図24に示す話題対に絞り込むことができる。図24では、閲覧対象文書全体同士の話題対を除いて、7対の関連話題しか抽出されていないのにも関わらず、関連話題に含まれなかった話題は、ノード左1、ノード左11、およびノード左15のみである。これらの話題のうち、質問項目を含むのはノード左15のみで、残りは後続の話題を導入するための役割を担った、答弁とは直接的に関連しない内容の部分であった。

【0127】

また、抽出された7対の関連話題は、後で示す結果に見られる通り、いずれも適切に対応している内容を含む部分であった。このように、本実施形態によれば、話題階層を利用して関連話題を絞り込むことで、あらかじめ特別な閾値を設定しなくても、過不足なく適切な話題対を選択することができる。

【0128】

次に、出力部28は、関連話題抽出部27が抽出した話題対のそれぞれについて、関連話題に対応する関連箇所をそれぞれの閲覧対象文書から切り出し、出力する。例えば、図24の（ノード左9-10、ノード右4-5）の関連度0.30の話題対については、ノード左9-10の話題に対応して、第1の閲覧対象文書中の区画9と区画10を、ノード右4-5の話題に対応して、第2の閲覧対象文書中の区画4と区画5を、それぞれ抽出し、利用者が対比しやすい形に整形して出力する。

【0129】

図25は、この話題対に対する関連箇所の出力結果の例を示している。図25の出力例では、左コラムが第1の閲覧対象文書の関連箇所に対応し、右コラムが第2の閲覧対象文書の関連箇所に対応している。それぞれの関連箇所は、話題階

層認定部 25 により認定された最小の話題（最小区画）の単位に分けて、出力されている。また、太字で強調出力されている語は、両方の関連箇所に出現し、かつ、それぞれの関連箇所において（4）式により求めた重要度が比較的大きい語である。具体的には、これらの語は、以下のような手順で抽出される。

【0130】

まず、両方の関連箇所に現れた語を重要語候補として抽出し、抽出された各語についてそれぞれの箇所における（4）式の値を、各語の各箇所における重要度として求める。そして、それぞれの箇所について、重要度の大きい順に重要語を抽出し、抽出した重要語における重要度の累積値が、重要語候補全体における重要度の合計値の $1/2$ を超えた時点で、抽出処理を打ち切る。

【0131】

なお、図 25 に示した関連話題は、本実施形態で用いた閲覧対象文書対において、重複して関連話題が抽出された唯一の部分である。この部分では、上位層の（ノード左 9-10，ノード右 4-5）の話題対だけでなく、それを構成する話題に関する（ノード左 9，ノード右 4）と（ノード左 10，ノード右 5）の話題対も関連話題として抽出されている。

【0132】

図 25 に示された内容に見られるように、ノード左 9 の質問に対する答弁はノード右 4 であり、ノード左 10 の質問に対する答弁はノード右 5 であると考えられるが、ノード左 9 とノード左 10 との間、および、ノード右 4 とノード右 5 との間にも、強い関連性が読み取れる。

【0133】

したがって、これらの関連話題が重複して抽出されているが、単に冗長なわけではなく、2 ノードずつのまとまり間の関係も、個々のノード間の関係も、ともに重要な意味を持っていることが分かる。そこで、図 25 では、まとまり全体を対比するだけでなく、個々の話題も対比できるように、対応する話題の開始位置を揃えて出力している。

【0134】

また、出力部 28 は、関連箇所の内容を要約して提示することで、関連箇所の

一覧性を向上させることもできる。例えば、上述した特開平 1 1 - 2 7 2 6 9 9 の技術を用いれば、上記の手順で抽出した重要語を多く含み、かつ、簡潔な要約を作成することができる。

【 0 1 3 5 】

図 2 6 は、このような要約処理の手順を簡略化して示したフローチャートである。出力部 2 8 は、まず、関連話題対に対応して、第 1 の閲覧対象文書から切り出された関連箇所 P 1 と第 2 の閲覧対象文書から切り出された関連箇所 P 2 を、入力として受け取る（ステップ S 1 2 1）。そして、関連箇所 P 1、P 2 のそれぞれから重要語を抽出し、それらの重要語をマージする（ステップ S 1 2 2）。

【 0 1 3 6 】

次に、関連箇所 P 1 から重要文を選択して要約を作成し（ステップ S 1 2 3）、関連箇所 P 2 から同様に要約を作成する（ステップ S 1 2 4）。そして、各関連箇所から作成された要約を比較しやすい形に整形し、対置して出力して（ステップ S 1 2 5）、処理を終了する。

【 0 1 3 7 】

図 2 7 は、図 2 6 のステップ S 1 2 3 および S 1 2 4 における重要文選択処理のフローチャートである。この処理では、出力部 2 8 は、まず、P 1 または P 2 を重要文の選択対象箇所 P に設定し、ステップ S 1 2 2 で抽出された重要語を、重要文の手掛かりとして重要語リスト K W L に設定する（ステップ S 1 3 1）。そして、対象箇所 P から重要語を最も多く含む文を重要文として選択し（ステップ S 1 3 2）、そのような文が選択できたか否かを判定する（ステップ S 1 3 3）。

【 0 1 3 8 】

文が選択できた場合は、次に、選択した文に含まれる重要語を重要語リスト K W L から取り除き（ステップ S 1 3 4）、K W L が空か否かをチェックする（ステップ S 1 3 5）。K W L が空でなければ、ステップ S 1 3 2 以降の処理を繰り返す。そして、すべての重要語に対して少なくとも 1 文以上の重要文が選択できた時点で選択処理を終了し、選択した文を元の文書における出現順に並べて、要約として出力して（ステップ S 1 3 6）、処理を終了する。

【 0 1 3 9 】

また、ステップ S 1 3 3 において、重要語を含む文がまったく選択できない場合は、処理を打ち切り、ステップ S 1 3 6 の処理を行う。図 2 6 および図 2 7 に示した処理により、図 2 8、2 9、および 3 0 に示すような要約が作成される。

【 0 1 4 0 】

このように、個々の関連話題に対応する関連箇所を個別に提示するだけでなく、要約機能により抽出された関連話題を短くまとめることで、関連箇所全体を一覧しやすい形で出力することができる。したがって、長い文書の比較・閲覧の際等に多くの関連話題が抽出されても、比較・閲覧作業を効果的に支援できる。

【 0 1 4 1 】

また、出力部 2 8 は、関連箇所と原文書である閲覧対象文書の全文とを対置して提示することで、閲覧対象文書における関連箇所の位置付け等を吟味しながら、関連箇所を分析する作業等を支援することができる。この場合、例えば、図 3 1 に示すような形態で、関連箇所の要約と閲覧対象文書の全文とを対置して提示すればよい。また、関連箇所と閲覧対象文書中の対応箇所との間にハイパーリンクを設定すれば、より一層閲覧効率を高めることができる。

【 0 1 4 2 】

図 3 1 では、左フレームが関連箇所参照のウィンドウであり、右フレームが閲覧対象文書参照用のウィンドウである。ここでは、左フレームに、作成された関連箇所の要約が表示されており、発言者名の後ろのかぎ括弧内（下線部）には、閲覧対象文書の対象箇所へのハイパーリンクのアンカーが設定されている。そして、利用者が必要に応じてアンカーを指定することで、第 1 の閲覧対象文書の指定箇所が右上のウィンドウに表示され、第 2 の閲覧対象文書の指定箇所が右下のウィンドウに表示される。

【 0 1 4 3 】

右フレームに提示された閲覧対象文書では、関連箇所が下線付きで強調表示され、関連箇所とその前後の文脈とを区別できるようになっている。強調表示の方法としては、カラー表示、網かけ表示等を用いることもできる。この例では、左フレームに関連箇所の要約が表示されているが、代わりに、関連箇所そのものを

表示してもよい。また、利用者の要望に応じて、関連箇所を要約で提示するか、関連箇所の内容全体を提示するかを切り換えられるようにすることも考えられる。

【0144】

また、出力部28は、閲覧対象文書間の関連箇所の関係をグラフ等を用いて図面で表現することで、閲覧対象文書相互の全体的な関連性を一目で把握できるようにすることもできる。この場合、例えば、図32に示すような形態で、閲覧対象文書中の関連箇所の出現状況を提示すればよい。

【0145】

図32では、上フレームに、2つの閲覧対象文書の話題階層が、図24と同様のグラフ形式で表示されており、下フレームに、第1および第2の閲覧対象文書が対置して表示されている。話題階層のグラフには、関連話題を示すアークが付加されており、それぞれのアークには、両方の文書の関連箇所を同期して提示するためのハイパーリンクが設定されている。また、話題に対応する各ノードには、個々の閲覧対象文書の対応箇所へのハイパーリンクが設定されている。さらに、各閲覧対象文書には、図31と同様に、関連箇所に強調表示が施されている。

【0146】

このように、閲覧対象文書間の関連箇所の出現状況をグラフ等により提示すれば、閲覧対象文書間の全体的な関連性の理解が促進される。これにより、例えば、本実施形態の閲覧対象文書のように全体として1対1で対応するような文書であるのか、それとも、特定の部分のみに集中して関連箇所が現れる文書であるのかを、利用者が一目で読み取れるようになる。そして、後者のような文書であれば、利用者は、関連箇所が集中している部分に絞って、効率的に文書の閲覧を進めることも可能になる。

【0147】

以上の実施形態では、主として、閲覧対象文書が2つの場合について説明したが、この処理を応用して、3つ以上の文書に対する比較・閲覧を支援することも可能である。例えば、いずれか1つの文書を基準（軸）として他の文書を比較するか、または、すべての閲覧対象文書の対に対して上述のような処理を行い、抽

出された関連話題を何らかの方法で整理・統合して出力すればよい。後者の場合、関連話題の整理・統合の方法としては、例えば、いずれか1つの文書中の同一箇所に対応する他の文書的话题を統合することが考えられる。

【 0 1 4 8 】

例えば、上述の「第149回衆議院本会議会議録第2号」に含まれる全8件の代表質問（答弁は除外）を、それぞれ異なる文書として切り出し、「第149回衆議院本会議会議録第1号」（2000年7月28日）中の首相の所信表明演説を基準文書として比較したところ、上述の第1の閲覧対象文書（水島広子議員による代表質問）にも関連する関連箇所として、図33のような関連箇所が抽出された。

【 0 1 4 9 】

図33では、左コラムが基準文書の関連箇所の要約に対応し、中央のコラムが第1の閲覧対象文書の関連箇所の要約に対応し、右コラムが他の文書の関連箇所に対応している。ここでは、第1の閲覧対象文書に関連する箇所のみを例として示したが、他の質問者による代表質問についても、同様に、基準文書の適切な箇所と対応付けることが可能である。

【 0 1 5 0 】

さらに、このような関連箇所を、基準文書に併合して出力することもできる。これにより、例えば、「所信表明演説のポイントとそれに対する各党代表者の見解」のような、統合的な文書の作成を支援することができる。

【 0 1 5 1 】

図34は、このような文書統合処理のフローチャートである。文書閲覧装置は、まず、利用者からの指示等に基づいて、複数の閲覧対象文書の中から基準文書を選択し（ステップS141）、上述の処理手順に従って、基準文書と他の各文書の間に関連箇所を抽出する（ステップS142）。そして、出力部28は、抽出された関連箇所を基準文書における出現順にマージして、統合文書を作成し（ステップS143）、その文書を出力して（ステップS144）、処理を終了する。

【 0 1 5 2 】

次に、英語の文書に対する処理について、1999年のケルンサミットと2000年の沖縄サミットにおけるG8による2つのコミュニケを対象とした場合を例に挙げて説明する。ここでは、外1を第1の英語閲覧対象文書とし、外2を第2の英語閲覧対象文書とした。

【0153】

【外1】

“G8 COMMUNIQUÉ KÖLN 1999”

【0154】

【外2】

“G8 COMMUNIQUÉ OKINAWA 2000”

【0155】

これらの文書の全文は、それぞれ、4千5百語および7千語であり、本明細書および図面にすべての処理結果を記載するには長すぎるため、以下では全体の半分ほどの部分のみを対象として処理を行った。全体で10節からなる第1の閲覧対象文書では、以下の5節（1千8百語）を処理対象とし、第2の閲覧対象文書では、前置き（Preamble）の次に位置する以下の1パート（3千5百語）を処理対象とした。

（1）第1の閲覧対象文書の処理対象部分

- I. Getting the World Economy on Track for Sustained Growth
- II. Building a World Trading System That Works for Everyone
- III. Designing Policies for More Employment
- IV. Investing in People
- V. Strengthening Social Safeguards

(2) 第2の閲覧対象文書の処理対象部分

Toward a 21st century of greater prosperity

また、ここでは、次のような処理方法およびパラメータを採用した。

(1) 単語認定の方法：ストップワードリストを用いた方法

(2) 結束度計算用の窓の幅：

最小窓幅 $w_{min} = 80$ (語)

最大窓幅 w_{max} w_{min} の2の中乗倍で、文書全体の $1/2$ を超えない値の語数

刻み幅 窓幅の $1/8$

図35は、第1の閲覧対象文書の先頭部分を示しており、図36は、その部分に対する単語認定部22の処理結果を示している。図36において、[]で括られた部分が、認定された単語に対応する。先頭の1文字のみが大文字の単語は、[]内ではすべて小文字に置き換えられている。

【0156】

この場合、単語認定部22は、まず、空白および“、”、“.”、“:”、“;”等の区切り記号を手掛かりに単語を切り出し、次に、それらの単語のうち、図37に示すようなストップワードリストに含まれる単語を取り除く、という手順で単語認定を行う。ストップワードリストとは、重要語として抽出したくない冠詞、前置詞等の単語を、あらかじめ定義したリストである。

【0157】

図38は、上述の文書対に対する関連話題の抽出結果を示している。図38において、左の木構造グラフが、第1の英語閲覧対象文書に対する話題階層認定部25の出力、すなわち、第1の英語閲覧対象文書の話題階層の認定結果に対応し、右の木構造グラフが、第2の英語閲覧対象文書の話題階層の認定結果に対応する。また、これらの木構造のノード間のアークは、関連話題抽出部27が抽出した関連話題対を表す。

【0158】

こうして抽出された関連話題を、出力部28が図26および図27の手順により要約した結果、図39、40、および41に示すような要約が得られた。

このように、本発明は、英語の閲覧対象文書に対しても、日本語の文書と同様に適用可能であり、また、任意の言語および任意の形式の文書に対して適用され、同様の結果を得ることができる。

（付記 1） 閲覧対象として指定された複数の文書を利用者に提示する文書閲覧装置であって、

前記複数の文書のそれぞれの話題階層を認定する話題階層認定手段と、

認定された話題階層に基づき、前記複数の文書に共通して現れる関連話題を抽出する関連話題抽出手段と、

抽出された関連話題に対応する記述箇所を、該複数の文書のそれぞれから切り出して出力する関連話題提示手段と

を備えることを特徴とする文書閲覧装置。

（付記 2） 前記関連話題抽出手段は、前記認定された話題階層中の様々な粒度の話題からなる話題対について、該話題対の各話題に対応する記述箇所の語彙的類似性に基づき、該話題対の話題間の関連度を求め、話題の包含関係に基づいて設定された閾値以上の関連度を持つ話題対を、前記関連話題として抽出することを特徴とする付記 1 記載の文書閲覧装置。

（付記 3） 前記関連話題提示手段は、切り出された記述箇所を、前記複数の文書間の関連箇所として対置して提示することを特徴とする付記 1 記載の文書閲覧装置。

（付記 4） 前記関連話題提示手段は、前記関連箇所と元の文書とを対置・連動して提示することを特徴とする付記 3 記載の文書閲覧装置。

（付記 5） 前記関連話題提示手段は、前記関連箇所の内容を要約して提示することを特徴とする付記 3 記載の文書閲覧装置。

（付記 6） 前記関連話題提示手段は、前記関連箇所の要約と元の文書とを対置・連動して提示することを特徴とする付記 5 記載の文書閲覧装置。

（付記 7） 前記関連話題提示手段は、前記複数の文書に対応する複数の話題階層と、前記複数の関連話題に基づく、該複数の話題階層中の話題間の対応関係とを、図面を用いて提示し、該図面上における前記利用者の指示に応じて、該複数の文書中の対応箇所を連動して提示することを特徴とする付記 3 記載の文書閲覧

装置。

（付記 8） 前記関連話題提示手段は、前記複数の文書のうちの 1 つの文書を基準文書として、前記関連話題に対応する記述箇所を他の文書から該基準文書に取り込んで新たな統合文書を作成し、該統合文書を出力することを特徴とする付記 1 記載の文書閲覧装置。

（付記 9） 閲覧対象として指定された複数の文書を利用者に提示するコンピュータのためのプログラムを記録した記録媒体であって、該プログラムは、
前記複数の文書のそれぞれの話題階層を認定し、
認定された話題階層に基づき、前記複数の文書に共通して現れる関連話題を抽出し、

抽出された関連話題に対応する記述箇所を、該複数の文書のそれぞれから切り出して出力する

処理を前記コンピュータに実行させることを特徴とするコンピュータ読み取り可能な記録媒体。

（付記 1 0） 閲覧対象として指定された複数の文書を利用者に提示するコンピュータにプログラムを伝送する伝搬信号であって、該プログラムは、
前記複数の文書のそれぞれの話題階層を認定し、
認定された話題階層に基づき、前記複数の文書に共通して現れる関連話題を抽出し、

抽出された関連話題に対応する記述箇所を、該複数の文書のそれぞれから切り出して出力する

処理を前記コンピュータに実行させることを特徴とする伝搬信号。

（付記 1 1） 閲覧対象として指定された複数の文書を利用者に提示する文書提示方法であって、

前記複数の文書のそれぞれの話題階層を認定し、

認定された話題階層に基づき、前記複数の文書に共通して現れる関連話題を抽出し、

抽出された関連話題に対応する記述箇所を、該複数の文書のそれぞれから切り出して出力する

ことを特徴とする文書提示方法。

【 0 1 5 9 】

【発明の効果】

本発明によれば、個々の閲覧対象文書の話題階層を用いて、複数の閲覧対象文書中の様々な粒度の話題が比較できるため、文書間で記述量の大きく異なる関連話題を適切に抽出することができる。また、抽出された関連話題に対応する関連箇所をそれぞれの閲覧対象文書から切り出し、それらを対置して出力することで、関連箇所を容易に比較・分析することが可能になる。このように、本発明によれば、複数の文書の比較閲覧作業が効果的に支援される。

【図面の簡単な説明】

【図 1】

本発明の文書閲覧装置の原理図である。

【図 2】

本発明の文書閲覧装置の構成図である。

【図 3】

情報処理装置の構成図である。

【図 4】

記録媒体を示す図である。

【図 5】

閲覧対象文書を示す図である。

【図 6】

単語認定処理のフローチャートである。

【図 7】

単語認定結果の例を示す図である。

【図 8】

形態素解析処理のフローチャートである。

【図 9】

日本語の辞書引きの例を示す図である。

【図 1 0】

英語の辞書引きの例を示す図である。

【図 1 1】

話題階層認定処理のフローチャートである。

【図 1 2】

結束度の系列を示す図である。

【図 1 3】

結束度分布の例を示す図である。

【図 1 4】

移動平均値と文書領域との関係を示す図である。

【図 1 5】

話題境界候補区間認定処理のフローチャートである。

【図 1 6】

結束力拮抗点を示す図である。

【図 1 7】

話題境界認定処理のフローチャートである。

【図 1 8】

関連付け対象データを示す図である。

【図 1 9】

話題境界の認定結果を示す図である。

【図 2 0】

第 1 の閲覧対象文書の話題階層を示す図である。

【図 2 1】

第 2 の閲覧対象文書の話題階層を示す図である。

【図 2 2】

関連話題抽出処理のフローチャートである。

【図 2 3】

関連度の計算結果を示す図である。

【図 2 4】

関連話題の抽出結果を示す図である。

【図 2 5】

関連箇所の出力例を示す図である。

【図 2 6】

要約処理のフローチャートである。

【図 2 7】

重要文選択処理のフローチャートである。

【図 2 8】

関連箇所の要約例を示す図（その 1）である。

【図 2 9】

関連箇所の要約例を示す図（その 2）である。

【図 3 0】

関連箇所の要約例を示す図（その 3）である。

【図 3 1】

原文書参照機能付き関連箇所提示の例を示す図である。

【図 3 2】

グラフ付き関連箇所提示の例を示す図である。

【図 3 3】

3 文書に対する関連箇所の出力例を示す図である。

【図 3 4】

文書統合処理のフローチャートである。

【図 3 5】

英語閲覧対象文書の先頭部分を示す図である。

【図 3 6】

英語閲覧対象文書の単語認定結果を示す図である。

【図 3 7】

ストップワードの例を示す図である。

【図 3 8】

英語の関連話題の抽出結果を示す図である。

【図 3 9】

英語の関連箇所の要約例を示す図（その１）である。

【図 4 0】

英語の関連箇所の要約例を示す図（その２）である。

【図 4 1】

英語の関連箇所の要約例を示す図（その３）である。

【符号の説明】

- 1 話題階層認定手段
- 2 関連話題抽出手段
- 3 関連話題提示手段
- 1 1 閲覧対象文書
- 1 2 文書閲覧装置
- 1 3 利用者
- 2 1 入力部
- 2 2 単語認定部
- 2 3 形態素解析部
- 2 4 単語辞書
- 2 5 話題階層認定部
- 2 6 話題境界候補区間認定部
- 2 7 関連話題抽出部
- 2 8 出力部
- 4 1 出力装置
- 4 2 入力装置
- 4 3 C P U
- 4 4 ネットワーク接続装置
- 4 5 媒体駆動装置
- 4 6 補助記憶装置
- 4 7 主記憶
- 4 8 バス
- 4 9 可搬記録媒体

5 0 サーバ

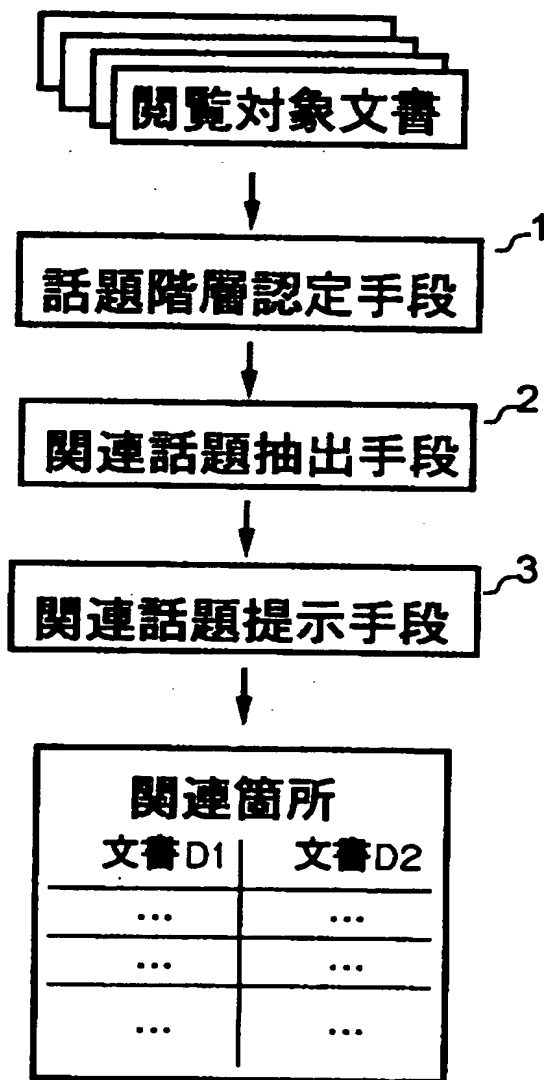
5 1 データベース

【書類名】

図面

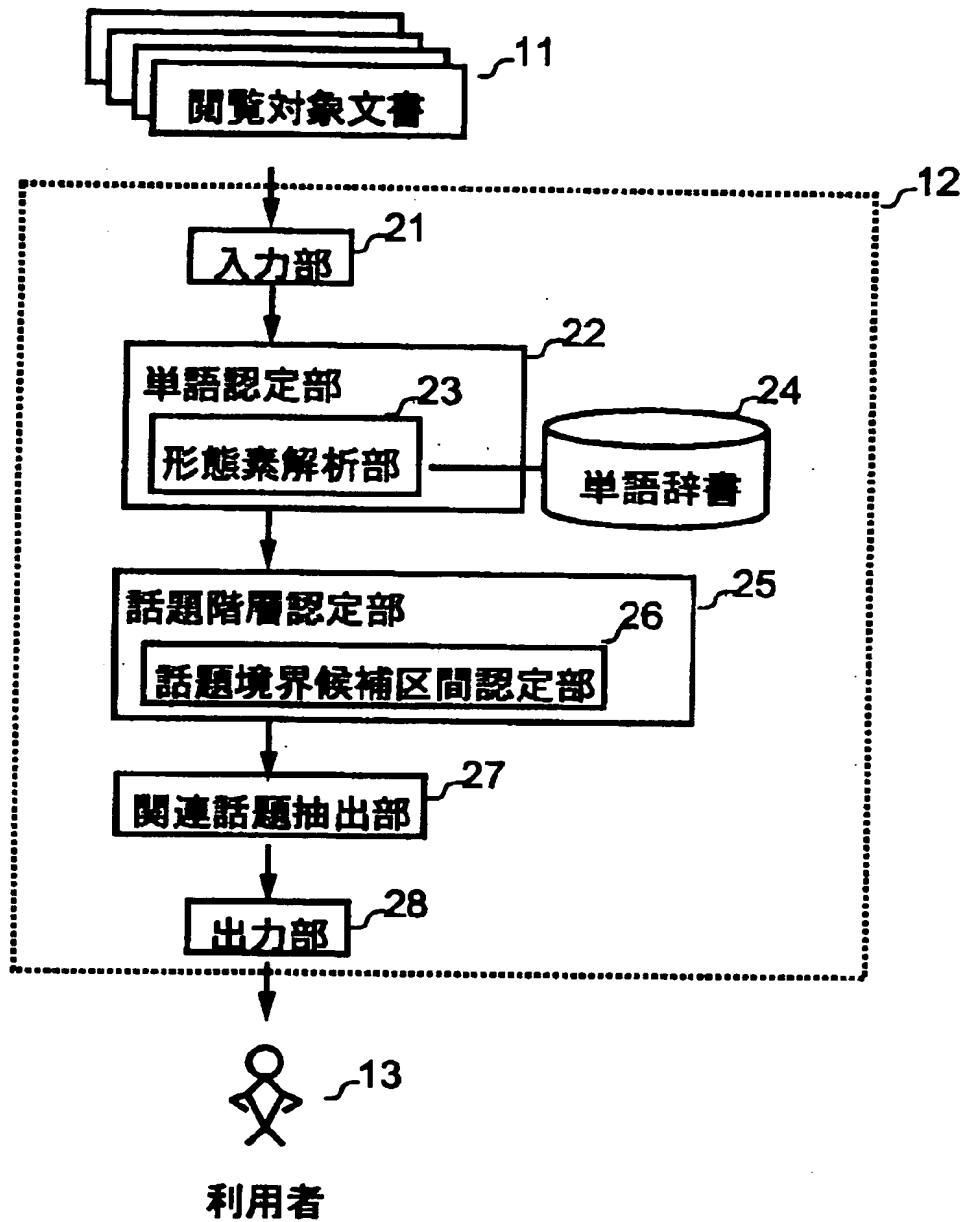
【図 1】

本発明の原理図



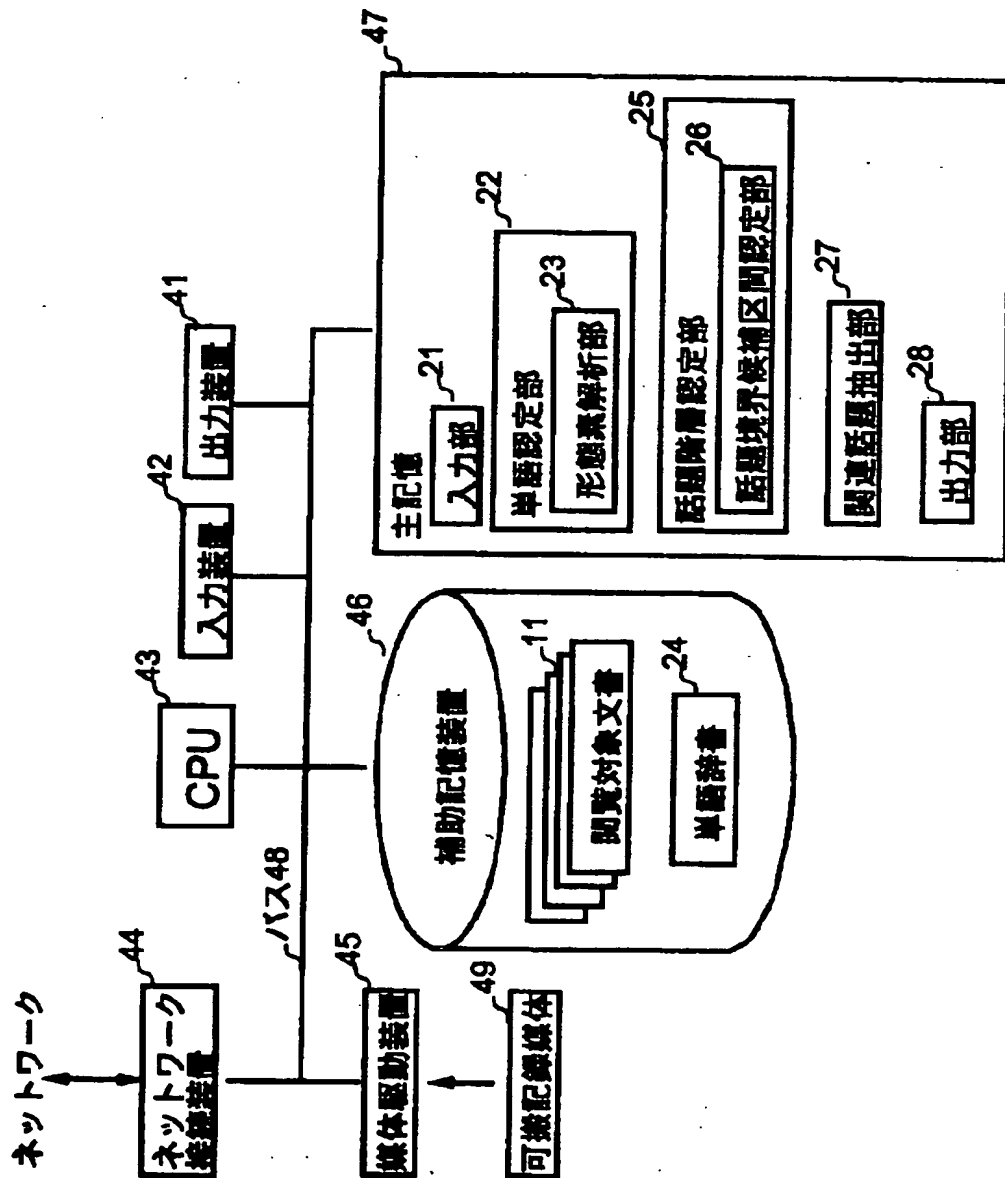
【図 2】

文書閲覧装置の構成図



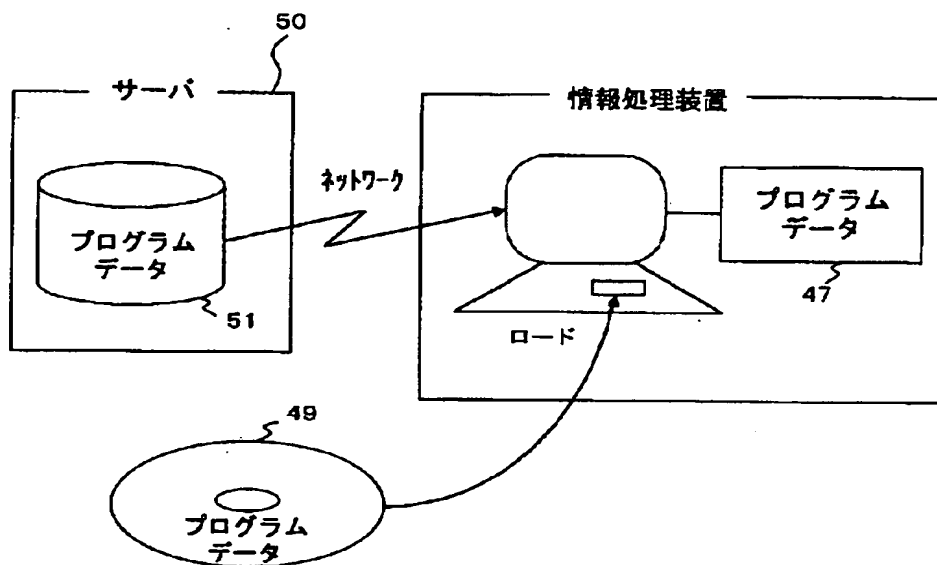
【図 3】

情報処理装置の構成図



【図4】

記録媒体を示す図



【図5】

閲覧対象文書の例を示す図

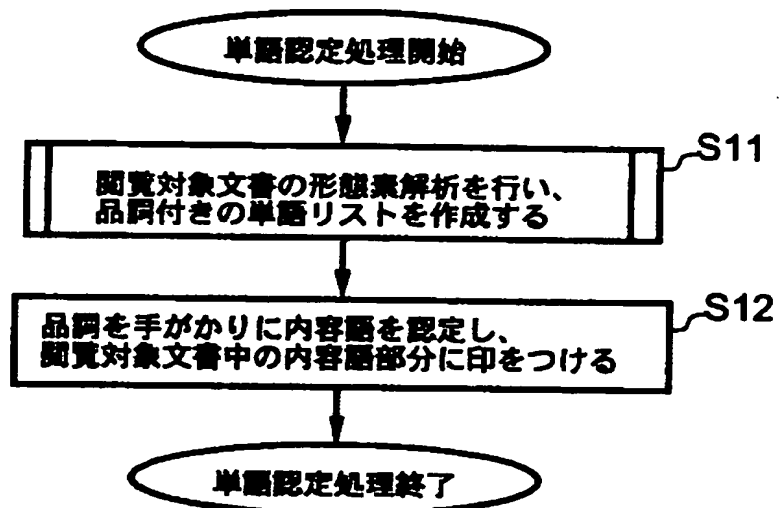
○水島広子君 私は、民主党・無所属クラブを代表して、森総理の所信表明演説に対し質問いたします。

初めに、有珠山や伊豆諸島における噴火、地震により亡くなられた方の御冥福をお祈りするとともに、被災者の皆様に心よりお見舞いを申し上げます。

今の日本の社会において子供の問題がかなり深刻であるということは、私も総理と同じ見解を持っております。私は、精神科医として、問題行動を起こす子供たちを数多く治療してきた経験から、日本の将来に非常な危機感を感じております。それが、精神科医である私が政治家を目指した最大の動機でもありました。少年犯罪についても、加害者に対する更生システムを専門化し徹底すると同時に、被害者のケアを充実するといった課題に目を向けずに、少年法を改正することで安易に厳罰化を図ろうとするような政治の姿勢には大きな危機を抱いております。(拍手)

【図 6】

単語認定処理のフローチャート



【図7】

単語認定結果の例を示す図

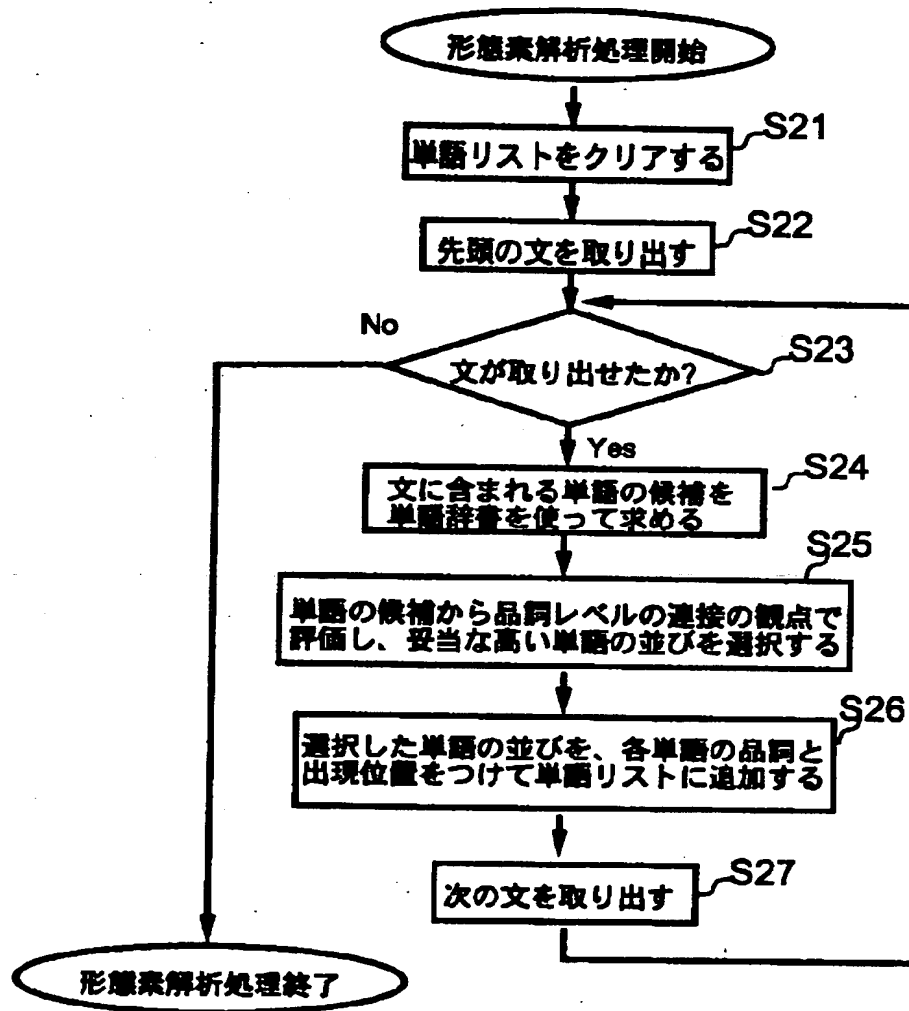
私は、【民主党/】・【無所属/】【クラブ/】を【代表/する】して、【森/】【総理/】の【所信/】
【表明/する】【演説/する】に【対/する】し【質問/する】いたします。

初めに、【有珠山/】や【伊豆/】【諸島/】における【噴火/する】、【地震/】に【よ/る】り【亡
くな/る】られた方の【御/する】【冥福/】を【お祈り/する】するとともに、【被災者/】の皆様
に【心/】よりお【見舞い/】を【申し上げ/る】ます。

今の【日本/】の【社会/】において【子供/】の【問題/】がかなり【深刻/】であると【い/
う】うことは、私も【総理/】と【同じ/】【見解/】を【持/つ】って【お/る】ります。私は、【精
神科医/】として、【問題/】【行動/する】を【起こ/す】す【子供/】たちを数多く【治療/する】
して【き/】た【経験/する】から、【日本/】の将来に非【常/】な【危機感/】を【感じ/る】て
【お/る】ります。それが、【精神科医/】である私が【政治家/】を【目指/す】した【最大/】の
【動機/】でもありました。【少年/】【犯罪/】についても、【加害者/】に【対/する】する【更生/
する】【システム/】を【専門化/する】し【徹底/する】すると同時に、【被害者/】の【ケア/】
を【充実/する】すると【い/う】った【課題/】に【目/】を【向け/る】ずに、【少年/】【法/】
を【改正/する】することで【安易/】に【厳罰化/】を【図/】【ろう/】とするような【政治/】
の【姿勢/】には大きな【危機/する】を【抱/く】いて【お/る】ります。

【図 8】

形態素解析処理のフローチャート



【図 9】

日本語の辞書引きの例を示す図

入力文	東京都は大都市だ	
	見出し (語幹)	品詞
候補単語	東	名詞
	東京都	名詞
	京都	名詞
	は	助詞「は」
	大	接頭語
	都市	名詞
	だ	助動詞「だ」

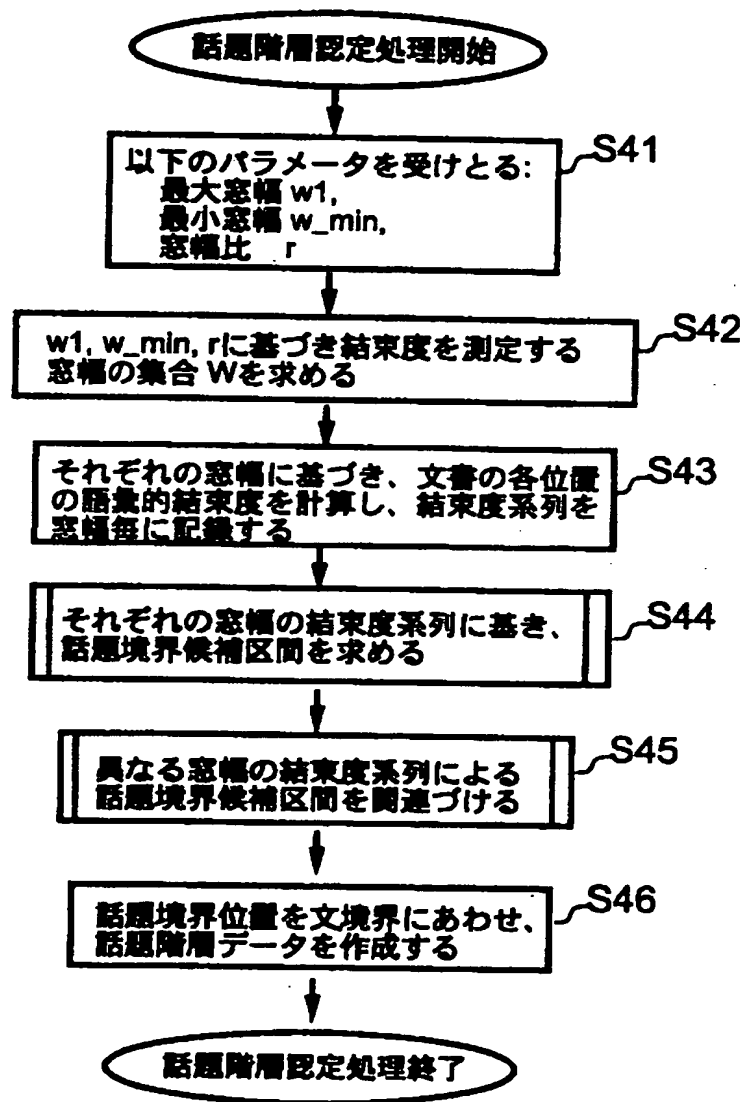
【図 10】

英語の辞書引きの例を示す図

入力文	Tokyo is the Japanese capital.		
	headword	base(root) form	part of speech
候補単語	Tokyo	Tokyo	proper noun
	is	be	be verb (the third person singular present form)
	the	the	definite article
	Japanese	Japanese	proper noun
	Japanese	Japanese	adjective
	capital	capital	noun

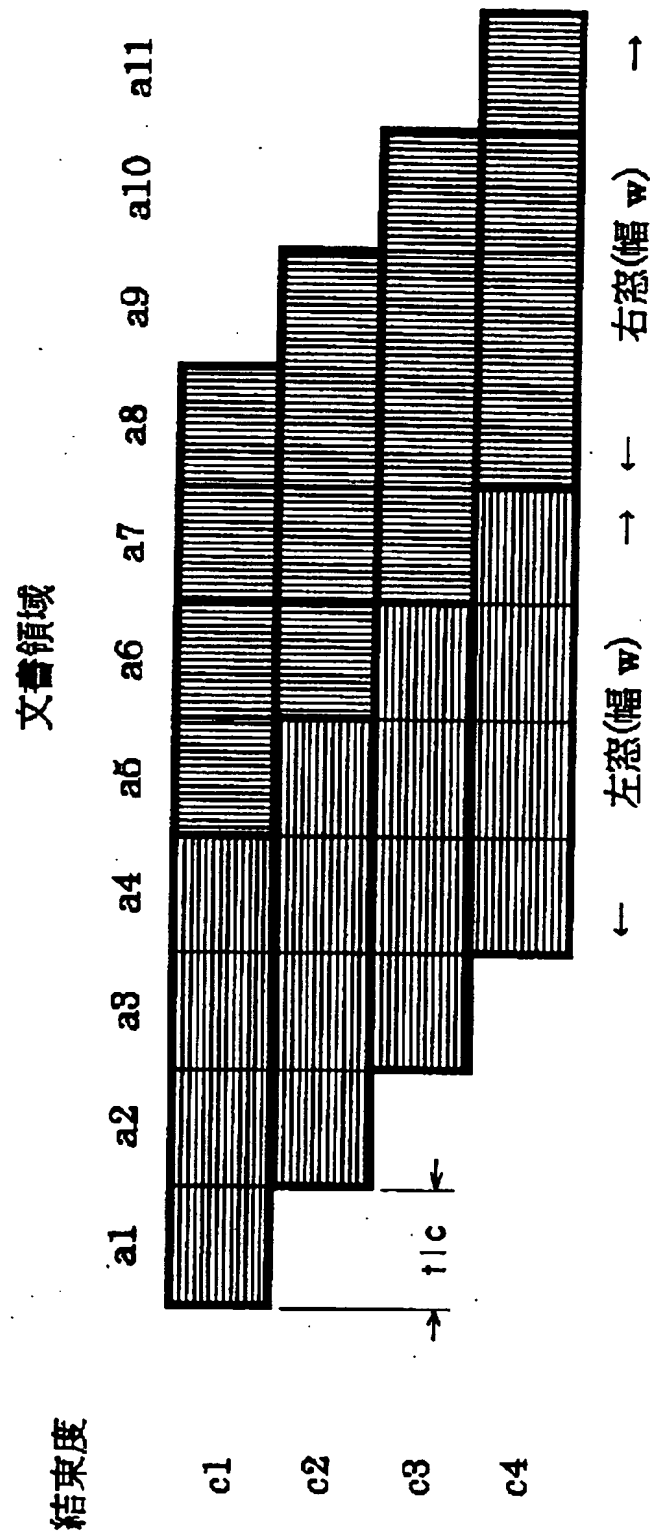
【図 1 1】

話題階層認定処理のフローチャート



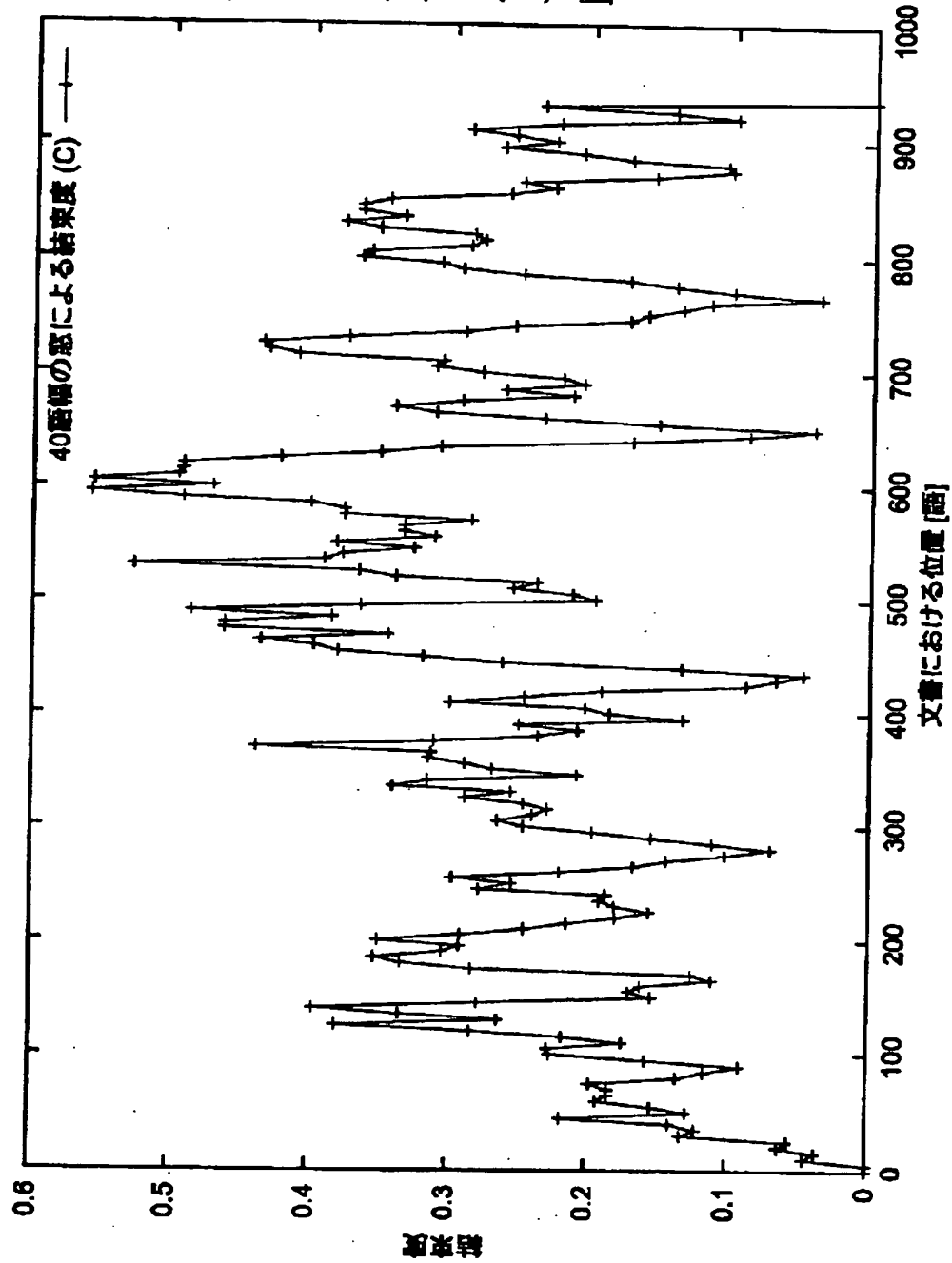
【図 1 2】

結束度の系列を示す図



【図13】

結束度分布の例を示す図



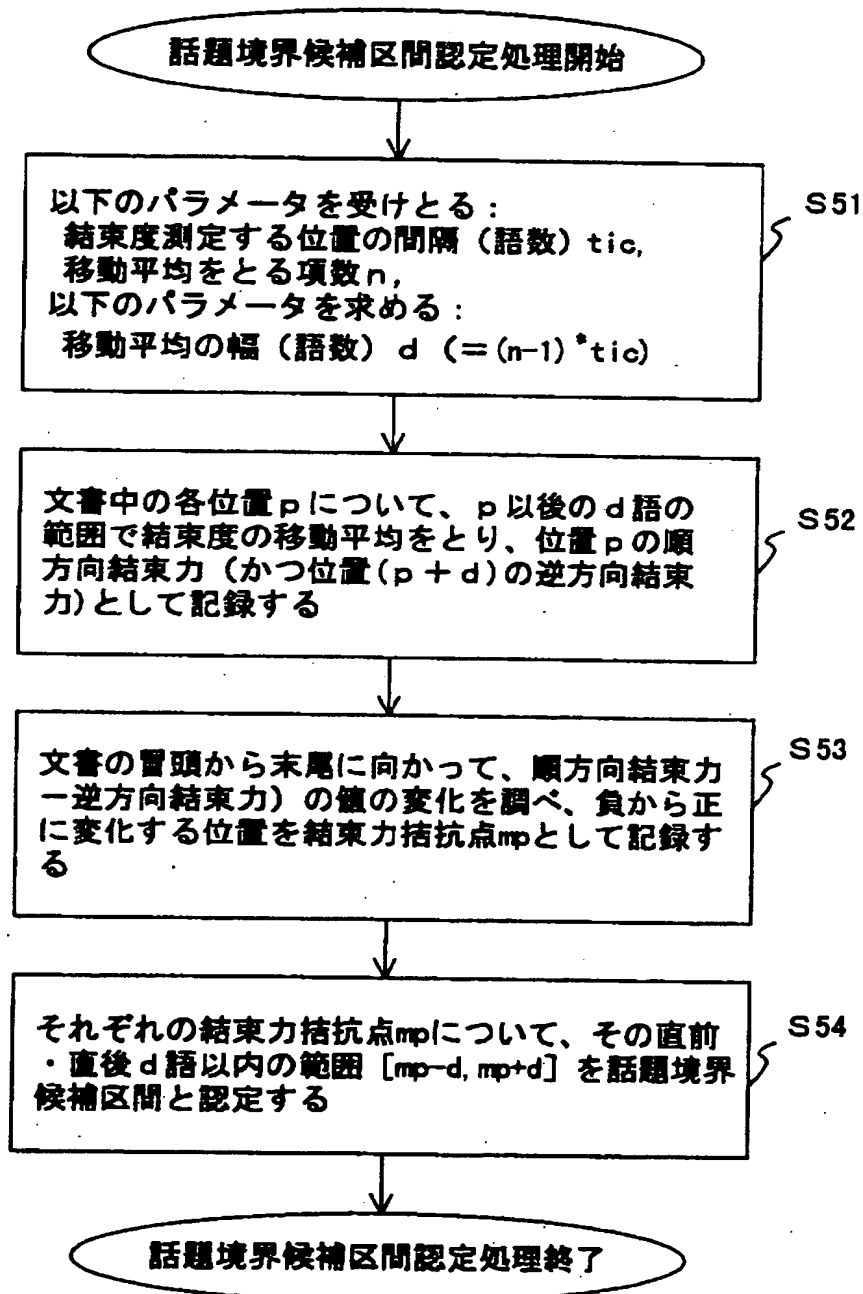
【図14】

移動平均値と文書領域との関係を示す図

項目数		文書領域と関係のある平均された結束度の数										
		a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	a11
4項平均 (c1~c4)	左窓	1	2	3	4	3	2	1	0	0	0	0
	右窓	0	0	0	0	1	2	3	4	3	2	1
3項平均 (c1~c3)	左窓	1	2	3	3	2	1	0	0	0	0	
	右窓	0	0	0	0	1	2	3	3	2	1	
2項平均 (c1, c2)	左窓	1	2	2	2	1	0	0	0	0		
	右窓	0	0	0	0	1	2	2	2	1		

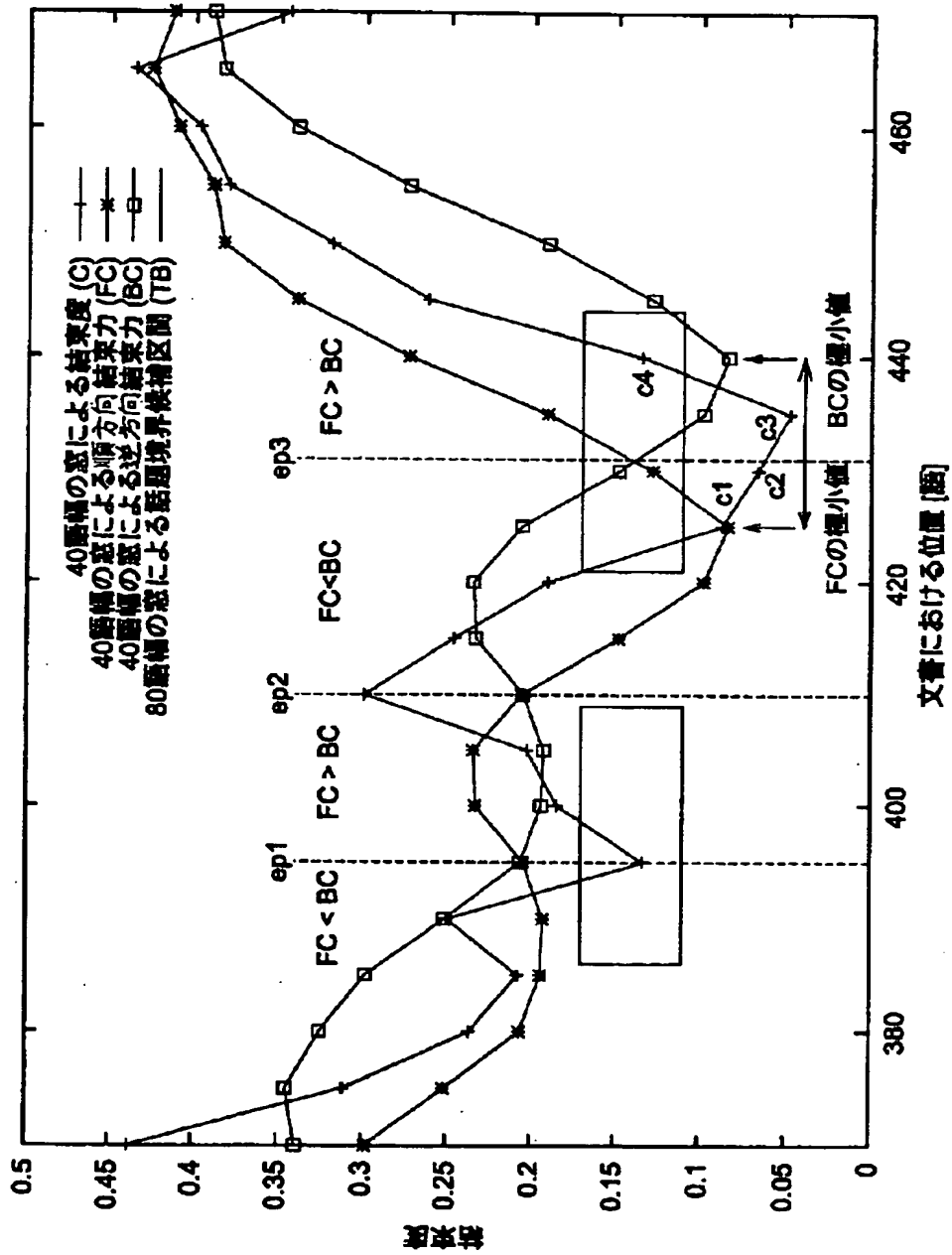
【図 15】

話題境界候補区間認定処理のフローチャート



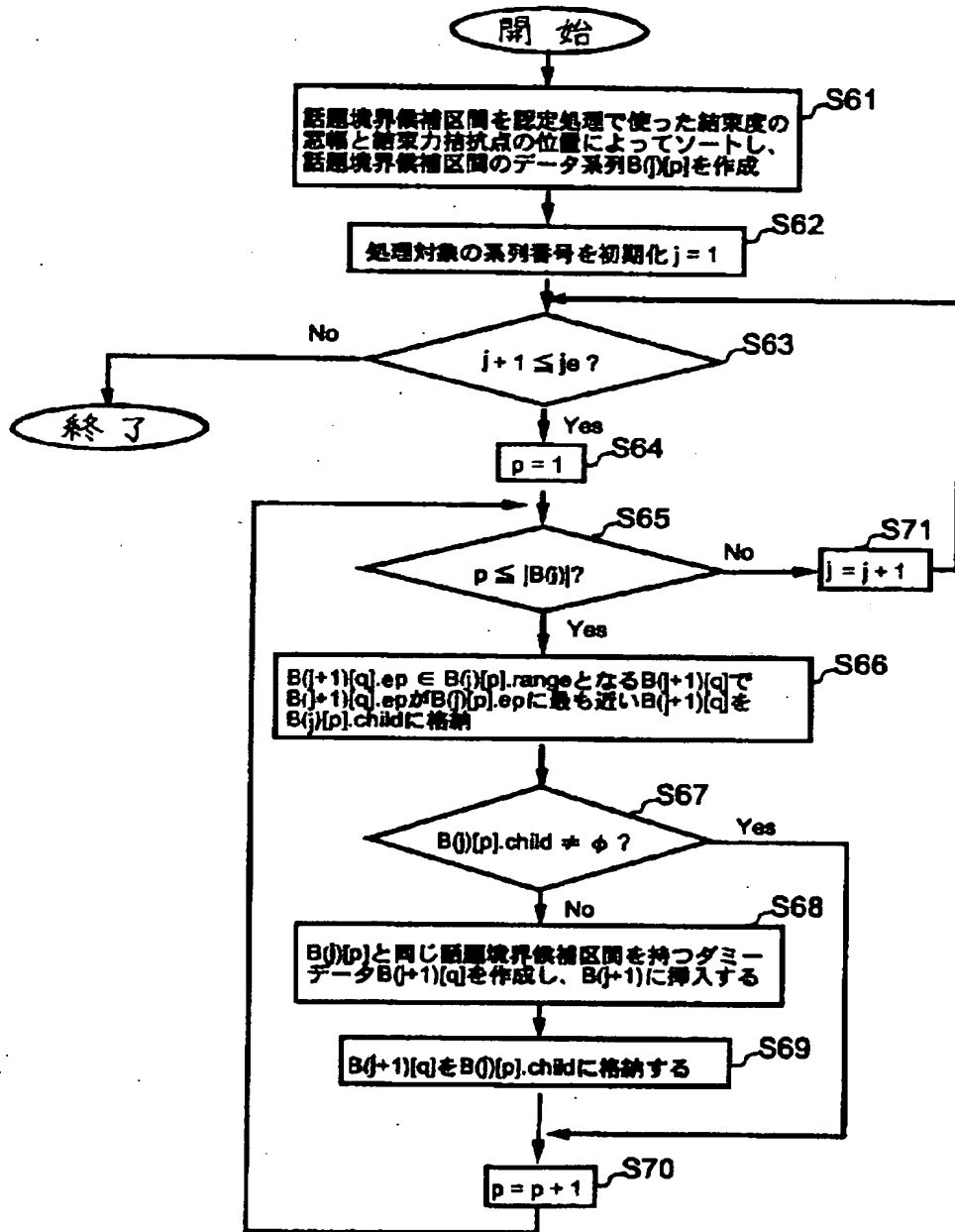
【図 16】

結束力拮抗点を示す図



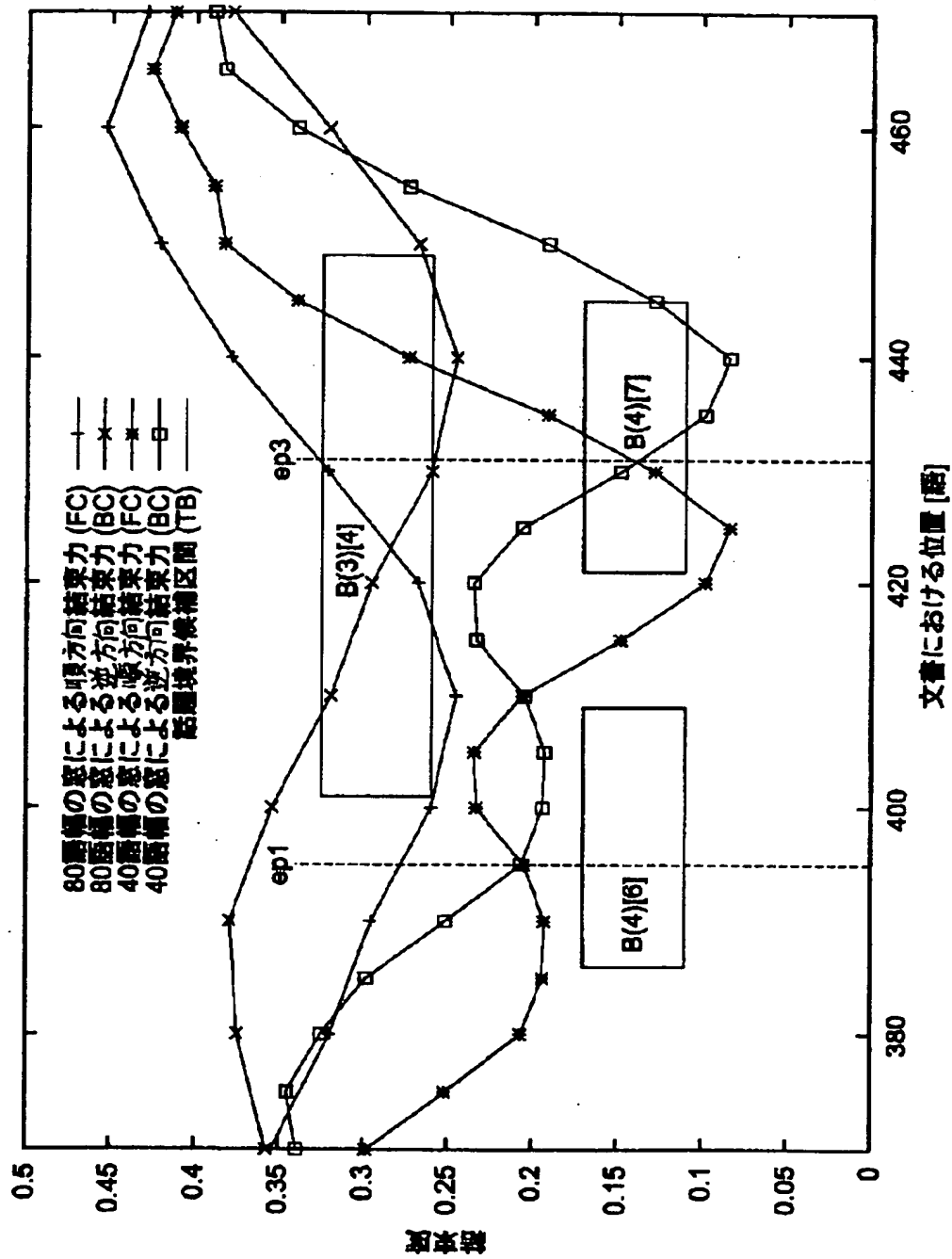
【図 17】

話題境界認定処理のフローチャート



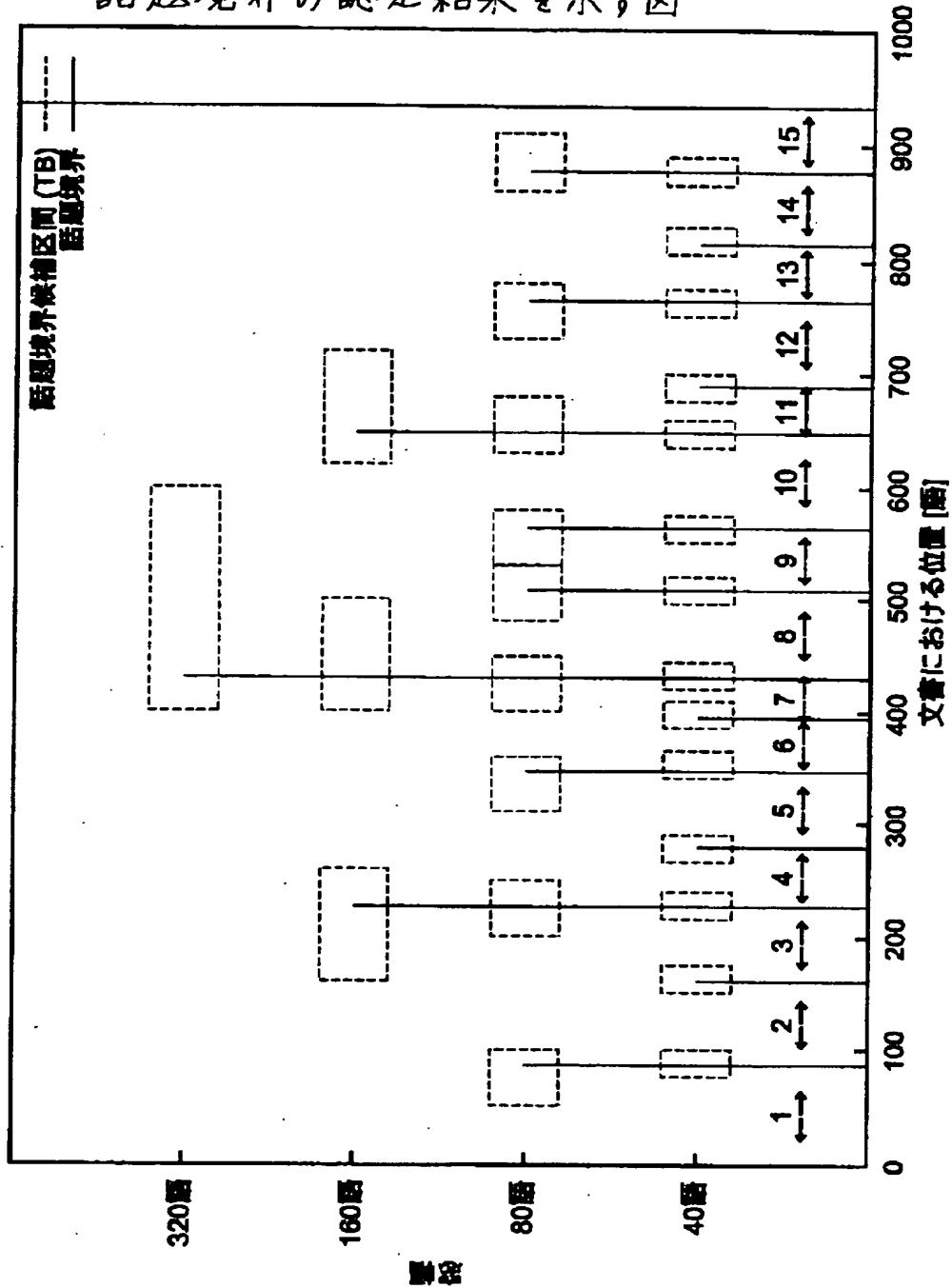
【図18】

関連付け対象データを示す図



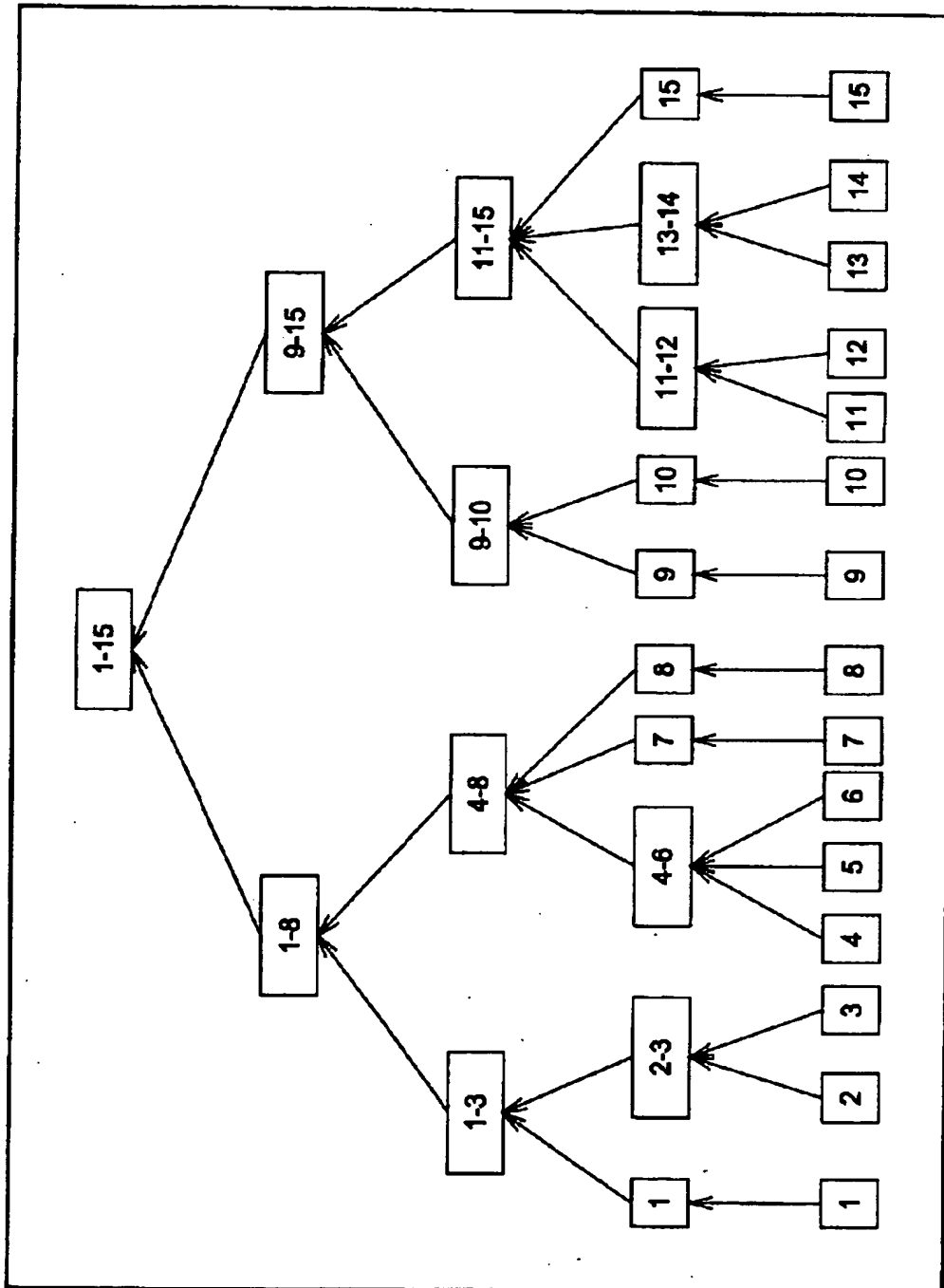
【図19】

話題境界の認定結果を示す図



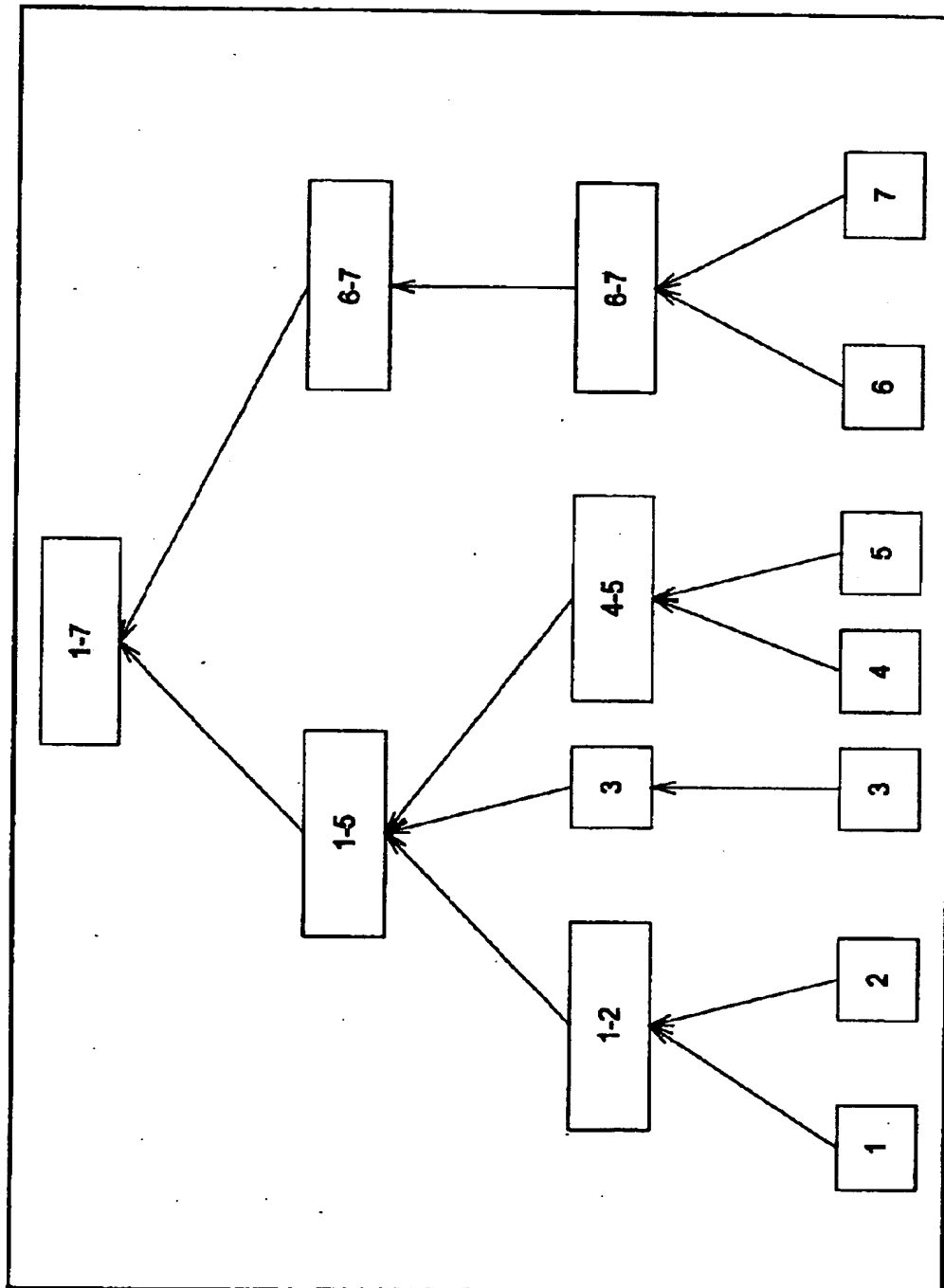
【図 20】

第 1 の閲覧対象文書の話題階層を示す図



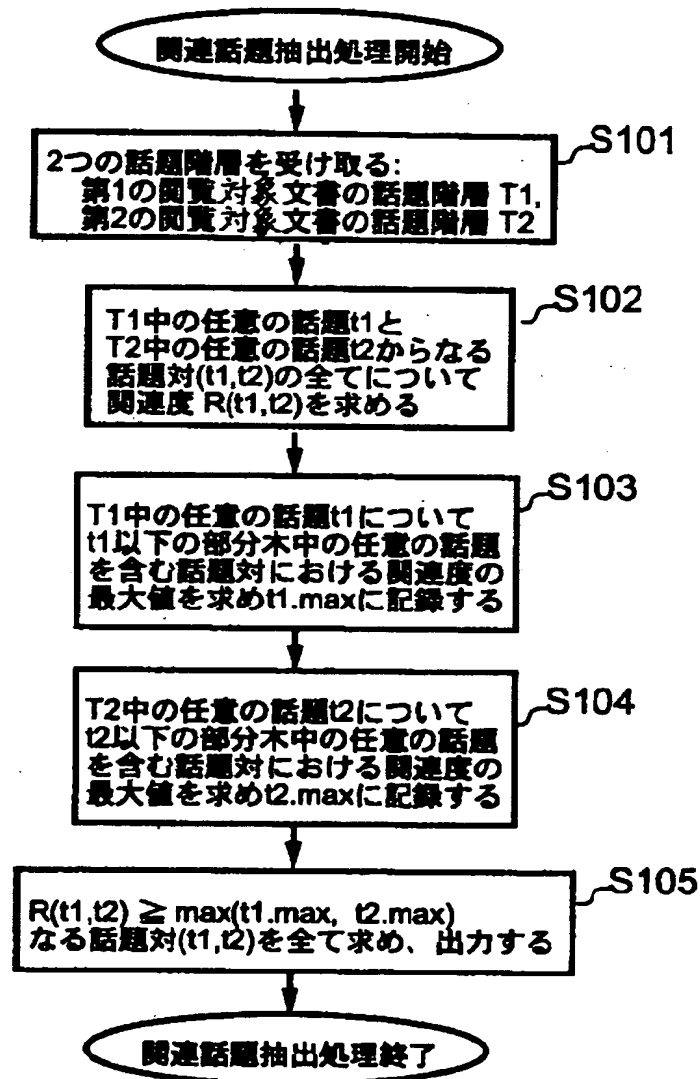
【図21】

第2の閲覧対象文書の話題階層を示す図



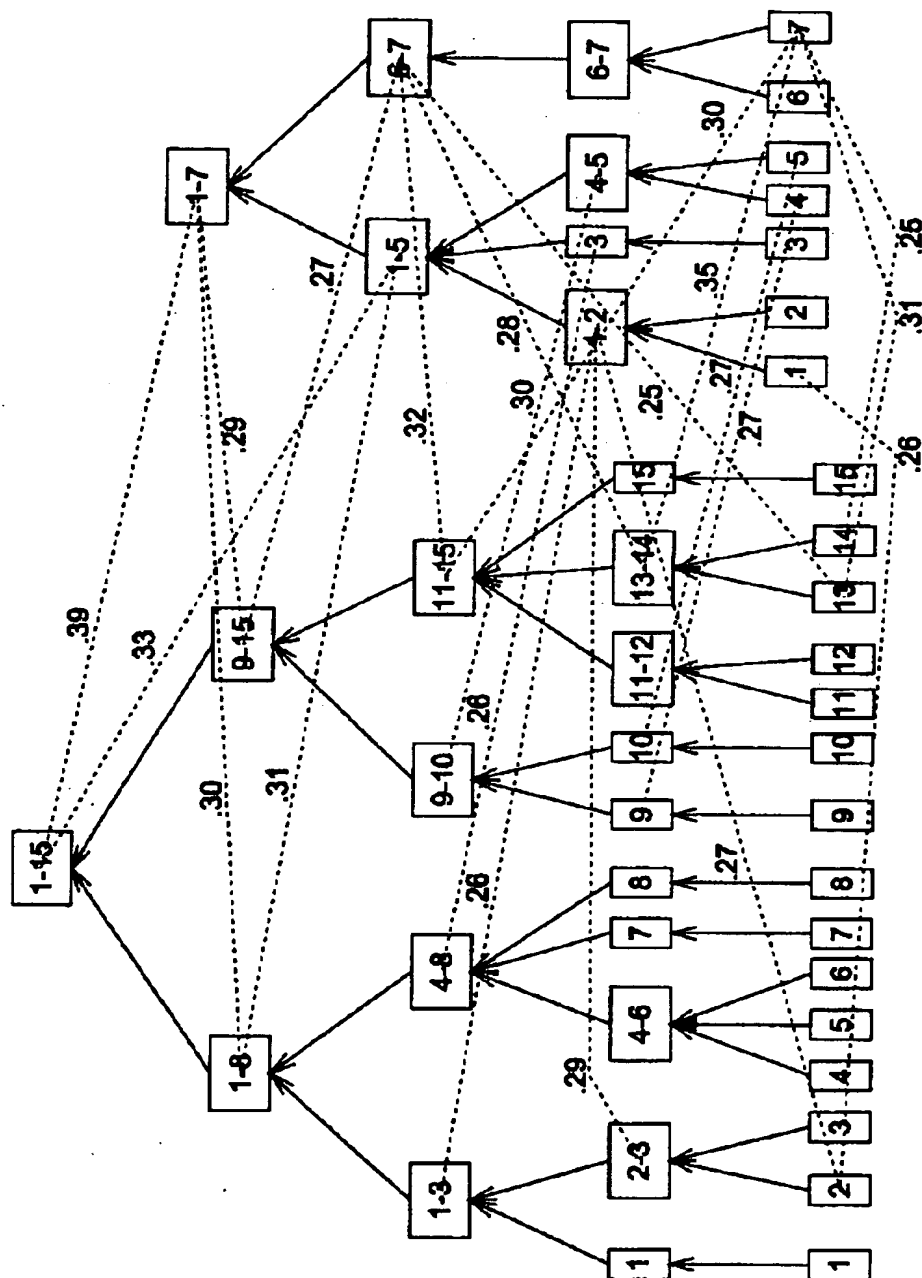
【図 22】

関連話題抽出処理のフローチャート



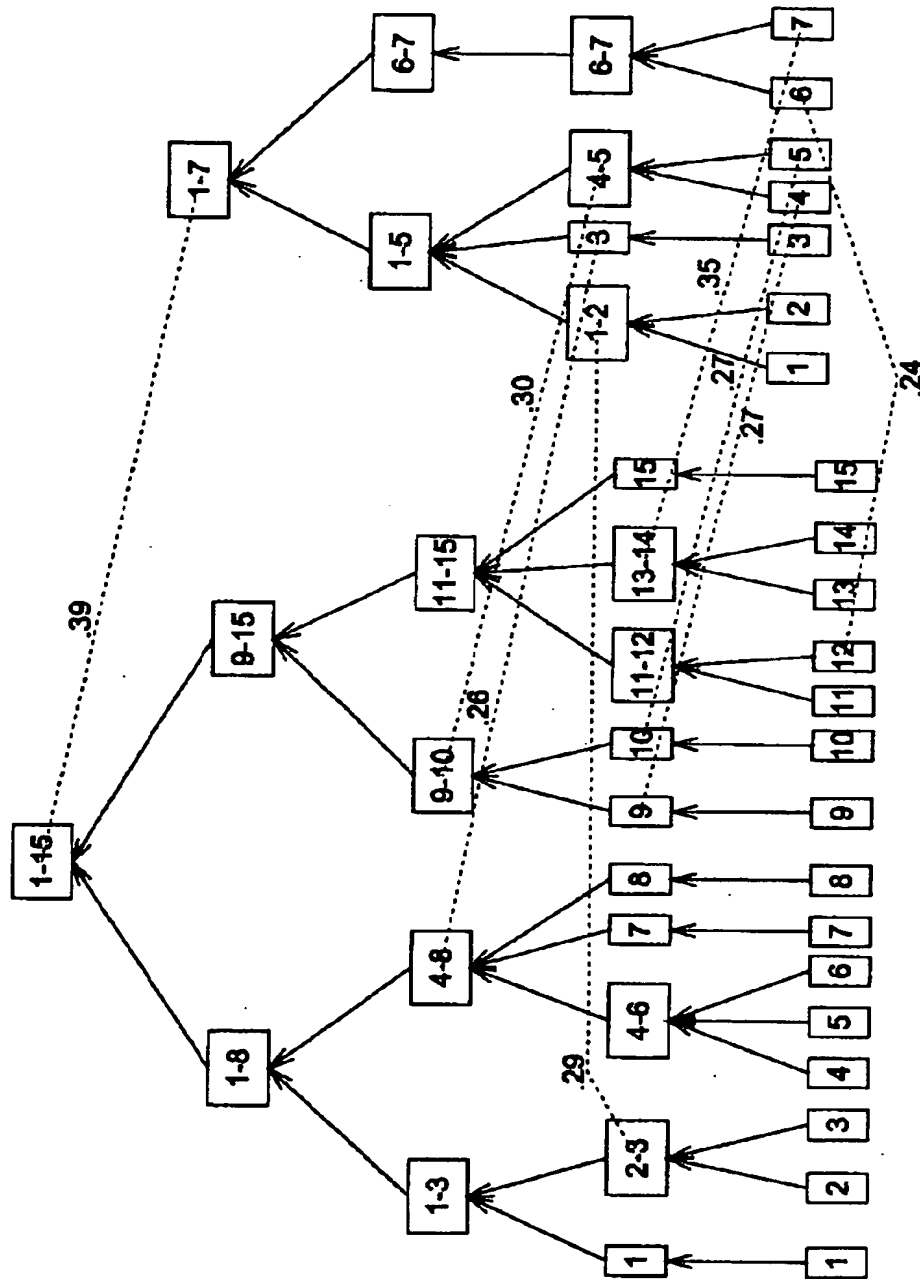
【図23】

関連度の計算結果を示す図



【図24】

関連話題の抽出結果を示す図



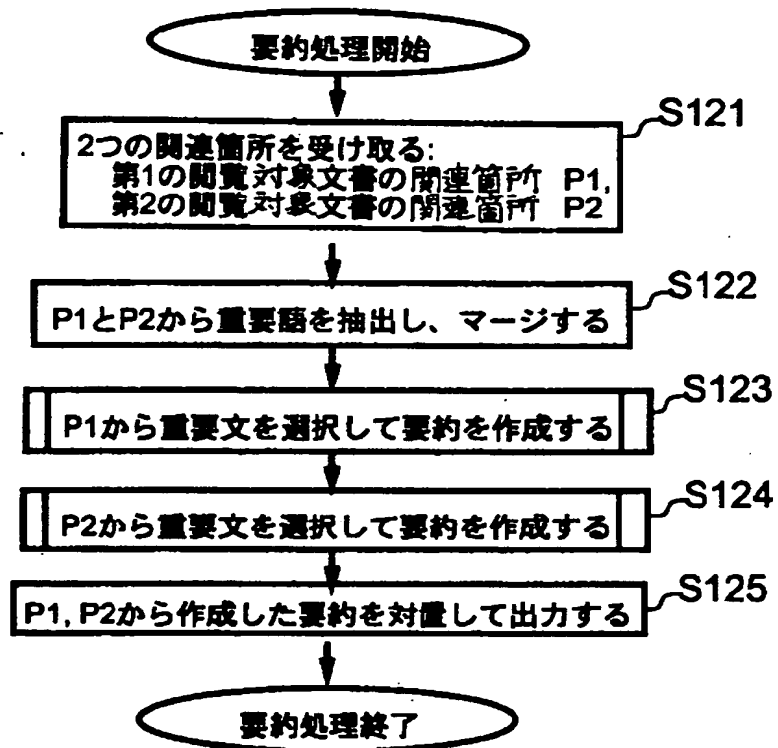
【図25】

関連箇所出力例を示す図

<p>○水島広子君 [9-10]</p> <p>⑨ 総理御自身も触れられている大人社会のあり方ですが、これが子供たちに大きな影響を与えるのは事実だと思います。子供たちは大人のまねをして成長します。大人社会のモラルがこれほど低下した今の日本で、子供たちのモラルだけが高まったら、むしろおかしいことだと思います。モラルの低下の一つの例として、子供の目に触れるテレビや雑誌、ゲームなどの影響も無視できません。だれでも簡単に目にするメディアに暴力や性暴力がはらんし、町じゅうに充塞情報があふれているというのが今の大人の社会です。子供たちを批判する前に、総理御自身も含めて、私たち大人がまず反省すべきではないでしょうか。</p> <p>⑩ 子供たちの問題行動とメディアによる有害情報の関係を指摘する専門家はたくさんいます。仮に犯罪に直結しなくても、幼いころから有害情報に当たり前のように触れることが子供たちの精神面の発育に及ぼす影響は無視できません。諸外国でも進められているように、子供たちを有害な情報から守る法律を日本でも早急につくる必要があると思います。これはもちろん、国家による検閲というような形をとるべきではありません。例えば、子供にとって有害な情報であるか否かを親が判断して選べるようなシステム、また、町中でも子供が有害情報に触れるのを防ぐような社会的なバリアをつくるなど、地域社会の大人たちが子供たちを守るようなシステムをつくるべきだと思います。子供を有害情報から守るための立法の必要性について、森総理はいかにお考えでしょうか。</p>	<p>○内閣総理大臣（森喜朗君） [4-5]</p> <p>④ テレビや雑誌、ゲームなどの青少年を取り巻く環境について、暴力や性犯罪がはらんしており、青少年にとって大きな問題であるとの御指摘であります。これらの問題は、申すまでもなく大人社会の責任であります。青少年を取り巻く社会環境の改善のため、社会が一体となった取り組みを進めることが極めて重要であると考えております。</p> <p>⑤ また、子供たちを有害情報から守るための法律の早急な制定を促す御意見をいただきました。私は、かねてから、少年非行対策は与野党対立案件にあらずと考えておりますが、御指摘の点については、まさに議員と意見を一にするものであります。しかしながら、この種法律の制定につきましても、青少年をめぐる環境の浄化の基本的なあり方や表現の自由とのかかわりなど、国民的な合意の形成が必要であると考えられ、関係方面の幅広い議論を重ねていきたいと考えております。</p>
---	--

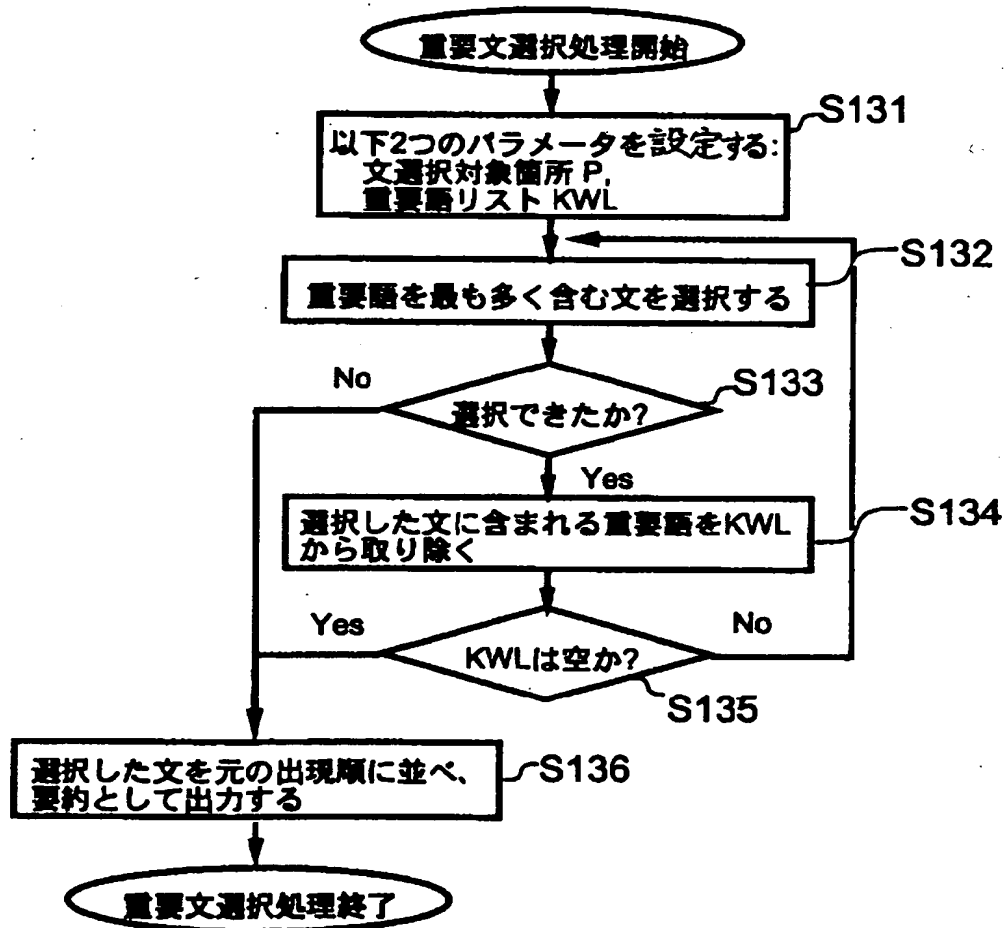
【図 2 6】

要約処理のフローチャート



【図 2 7】

重要文選択処理のフローチャート



【図28】

関連箇所の要約例を示す図 (その1)

<p>○水島広子君 [2-3]</p> <p>…精神科医として現場で子供たちや親たちと向き合ってきた私の目には、総理のおっしゃるような教育改革で問題が解決できるとはとても思えません。…いじめの問題を根本的に解決するには、人間の多様性を尊重して、自分も他人も大切にできる子供を育てる教育が不可欠です。…教育基本法の見直しにしても、本来、他者との触れ合いを通して自発的に育てるはずの事仕の精神や道徳心といったものを法改正によって一方的に押しつけようとするのであれば、逆効果となり、取り返しがつかないことになると思います。…</p>	<p>○内閣総理大臣（森喜朗君）[1-2]</p> <p>…いじめについてのお尋ねですが、いじめ問題を解決するためには、弱い者をいじめることは人間として絶対に許されないとの認識のもと、事仕活動や自然体験活動などを通じて、子供たちに、命を大切に、他人を思いやる心など、基本的な倫理観をはぐくむことが重要であると考えております。…教育基本法についてのお尋ねですが、教育基本法は、創定以来半世紀を経過しております。…今後、教育改革国民会議において、例えば我が国の文化や伝統を尊重する気持ちを養う観点や生涯学習時代を迎える観点、あるいは教育において家庭や地域が果たすべき役割といった観点を初め、さまざまな観点から幅広く議論を進めていただきたいと考えております。民法改正についてのお尋ねがありました。</p>
<p>○水島広子君 [4-8]</p> <p>…夫婦は皆同じ名字にするというのは、明治維新に西洋のまねをして導入された制度であり、日本独自の伝統とは何ら関係がありません。…男女共同参画社会の実現を所信表明演説でもうたわれた総理は、希望する夫婦には夫婦別姓を認めるよう民法を直ちに改正することについて、どのようにお考えでしょうか、お尋ねいたします。…非嫡出子は、法律上相続のときに差別を受けるだけでなく、普通に社会生活を送る上でも、就職や結婚の際に差別を受けています。…</p>	<p>○内閣総理大臣（森喜朗君）[3]</p> <p>…この導入についてのお話ですが、これは婚姻制度や家族のあり方とも関連する重要な問題でありまして、国民や関係各方面の意見が現在分かれている状況にありますので、国民各層の御意見を幅広く聞き、また、各方面における議論の推移も踏まえながら、適切に対処していく必要があるのではないかと考えております。また、嫡出でない子の法定相続分等についてであります。民法上、嫡出でない子と嫡出である子の相続分に差異が設けられている点の解消につきましても、選択的夫婦別氏の問題と同様に、国民や関係各方面の意見が分かれている状況にありますので、国民各界各層の御意見をこれも幅広く聞くなどいたしまして、適切に対処していく必要があると考えております。</p>

【図29】

関連箇所の要約例を示す図 (その2)

<p>○水島広子君 [9-10]</p> <p>¶⁹ 総理御自身も触れられている大人社会のあり方ですが、これが子供たちに大きな影響を与えるのは事実だと思います。…モラルの低下の一つの例として、子供の目に触れるテレビや雑誌、ゲームなどの影響も無視できません。だれでも簡単に目にするメディアに暴力や性暴力がはらんし、町じゅうに売春情報があふれているというのが今の大人の社会です。…</p> <p>¶¹⁰ …諸外国でも定められているように、子供たちを有害な情報から守る法律を日本でも早急につくる必要があると思います。…</p>	<p>○内閣総理大臣（森喜朗君）[4-5]</p> <p>¶⁴ テレビや雑誌、ゲームなどの青少年を取り巻く環境について、暴力や性犯罪がはらんしており、青少年にとって大きな問題であるとの御指摘がありますが、これらの問題は、申すまでもなく大人社会の責任であります。…</p> <p>¶⁵ また、子供たちを有害情報から守るための法律の早急な制定を促す御意見をいただきました。…</p>
--	---

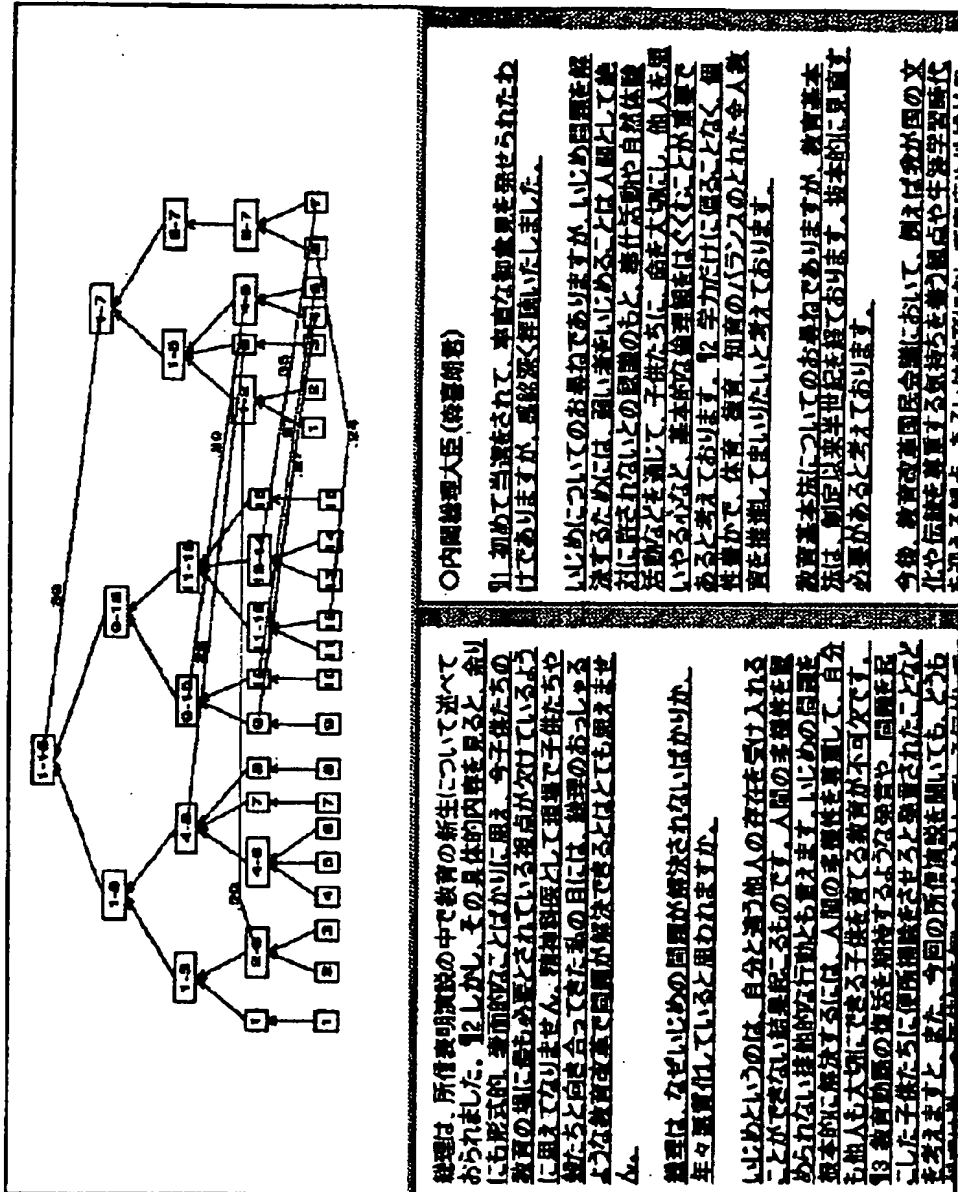
【図30】

関連箇所の要約例を示す図（その3）

<p>○水島広子君 [12]</p> <p>…しかし、最近の国会運営を見て、民主主義とは話し合いよりも多数決で押し切ることだという誤った理解をしている人たちがふえているように思います。…まずは党首討論を毎週行うことで総理みずから模範を示されるべきだと思いますし、あわせて、各大臣についても同様な場を設けるべきだと思いますが、総理のお考えをお願いいたします。</p>	<p>○内閣総理大臣（森喜朗君）[6]</p> <p>…積極的に議論を聞かせることは、最終的には多数決で決するにせよ、国民の前に争点を明らかにし、国民の政治への関心を高めるために重要なことであると考えます。お尋ねの党首討論のあり方などの国会の運営に関する問題につきましては、国会において御議論をいただきたいと思います。</p>
<p>○水島広子君 [13-14]</p> <p>…小児科救急の不備のために命を落とした不幸な子供の例も多数報道されています。また、心を病む子供に対しては、大人を中心とした医療体系では対応し切れません。小児の医療は、国の将来を担う貴重な人材を社会全体で守るという発想で行わなければいけないと思います。国が特別な予算枠を確保し良質な医療を提供すること、また、数少ない小児科医を効率的に活用するためにも、各都道府県に子供病院をつくり、人材を集中させ、包括的、専門的な子供医療ができるように国としても取り組むべきだと思いますが、いかがでしょうか。現行のように各都道府県の自助努力に任せていると、いまだに全国で十九都道府県にしか子供の総合医療施設がありません。…</p>	<p>○内閣総理大臣（森喜朗君）[7]</p> <p>…このため、小児専門の救急医療体制の整備、診療報酬における小児医療の適切な評価、小児医療施設の整備の補助などを行っているところでありますが、今後とも適切な対応をしていきたいと考えております。…</p>

【図32】

グラフ付き関連箇所提示を示す図



○内閣総理大臣(時憲閣答)

11 初めて当選をされて、早速な御意見を述べられました。けれども、私共は、御意見を深く拝聴いたしました。

いじめについてのお尋ねであります。いじめ問題を解決するためには、御意見をいじめることは人間として絶対に許されないと御指摘の通り、専ら活動や自然体験活動などを通じて、子供たちに、命を大切にし、他人を思いやる心など、基本的な倫理観をばくばくすることが重要であると考へております。12 学力だけに偏ることなく、個性がで、体育、道徳、知育のバランスのとれた全人教育を推進してまいりたいと考へております。

教育基本法についてのお尋ねであります。教育基本法は、創設以来半世紀を経ております。抜本的に見直す必要があると考へております。

今後、教育改革国民会議において、御見解が国の文化や伝統を尊重する気持を養う観点や生涯学習時代に

総理は、所得説明制度の中で教育の新生について述べておられました。12 しかし、その具体的な内容を見ると、余りにも形式的、表面的なことばかりに思え、今子供たちの教育の場に最も必要とされている視点が欠けているように思えてなりません。特に、子供たちと親で子供たちや親たちと向き合ってきた私共の目には、総理のおっしゃるような教育改革で問題が解決できるとはとて思えません。

総理は、なぜいじめの問題が解決されないばかりか、年々悪化していると思われませんか。

いじめというのは、自分と違う他人の存在を受け入れることができない結果起こるものです。人間の多様性を認められたい、排他的な行動とも重なります。いじめの問題を抜本的に解決するには、人間の多様性を尊重して、自分も他人も大切にできる子供を育てる教育が不可欠です。

13 教育活動の進捗を期するような検査や、同様に促した子供たちに便所掃除をさせると御指摘されたことなどを考えますと、また、今回の所信演説を聞いても、どうも

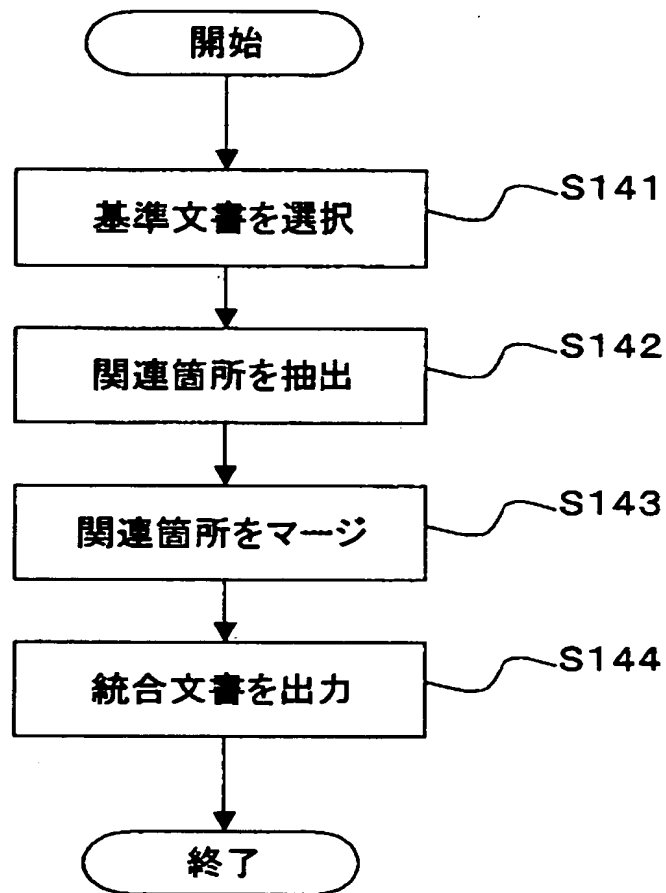
【図33】

3文書に対する関連箇所の出力例を示す図

<p>○内閣総理大臣（森喜朗君）[22-24] 日本新生プランの第三の柱は、教育の新生、すなわち教育改革であります。悪質な少年犯罪の続発や不登校、学級崩壊などの深刻化は、まことに心痛むものがあります。…命を大切にし、他人を思いやる心、事仕の精神、日本の文化、伝統を尊重し、国や地域を愛する気持ちをはぐくみ、二十一世紀の日本を支える子供たちが、創造性豊かな立派な人間として成長することこそが、心の豊かな美しい国家の礎と言えるのではないのでしょうか。私は、かねてから体育、徳育、知育のバランスのとれた全人教育を充実するとともに、世界に通用する技術、能力を備えた人材を育成するため、世界トップレベルの教育水準の確保が必要であると考えてきました。阪神・淡路大震災やナホトカ号重油流出事故のとき、全国津々浦々から若者たちが集まり、献身的にボランティア活動をしていた姿を見て、さすが日本の若者と感動したことを思い出します。…また、制定して半世紀となる教育基本法についても、抜本的に見直す必要があると考えております。教育改革国民会議においても、九月の中間報告に向けて、我が国の教育全般にわたり議論が行われているところであります。私は、学校の運営体制を整備するとともに、教師が、人間が人間を教えるというとうい使命感に燃えて教育に携わることが何よりも大切であり、IT教育や中高一貫の教育の推進、大学九月入学の推進、教員や学校の評価システムの導入、教育委員会のあり方なども重要な課題であると考えております。</p>	<p>○水島広子君 [1-3] …少年犯罪についても、加害者に対する更生システムを専門化し徹底すると同時に、被害者のケアを充実するといった課題に目を向けずに、少年法を改正することで安易に厳罰化を図ろうとするような政治の姿勢には大きな危惧を抱いております。総理は、所信表明演説の中で教育の新生について述べておられました。しかし、その具体的内容を見ると、余りにも形式的、表面的なことばかりに思え、今子供たちの教育の場に最も必要とされている視点が欠けているように思えてなりません。…いじめの問題を根本的に解決するには、人間の多様性を尊重して、自分も他人も大切にできる子供を育てる教育が不可欠です。…教育基本法の見直しにしても、本来、他者との触れ合いを通して自発的に育てるはずの事仕の精神や道徳心といったものを法改正によって一方的に押しつけようとするのであれば、逆効果となり、取り返しがつかないことになると思います。…</p>	<p>○土井たか子君 [17-22] …まさに全国的な課題と言ってよいでしょう。…むしろ、教育基本法改正や少年法改正の論議に深く踏み込むことによって問題は放置されてしまうと思えてなりません。…子供たちの荒れるのも、クラスの崩壊も、暴力も、子供たちが人間として尊重されていないところから始まっているように私には思えます。自分自身の命と人生を何よりも尊重され、大切に思える子供は、他人の命と人生も尊重し、大切にするものです。…事仕活動を学校で強制するなど、無意味どころか反発を呼んで逆効果であろうと私は思います。…ボランティアの強制など、言葉の矛盾であるばかりでなく、ようやく日本にも育ちつつある若者たちの本来のボランティア精神をも押し殺してしまうことになるでしょう。…総理、あなたは所信表明で、教育基本法を抜本的に見直す必要を言われました。…教育基本法の改正を提起しようとしている教育改革国民会議についてもお尋ねいたします。…</p>
---	---	--

【図 3 4】

文書統合処理のフローチャート



【図35】

英語閲覧対象文書の
先頭部分を示す図

I. Getting the World Economy on Track for Sustained Growth
4. Since we met last year in Birmingham, the world economy has faced major challenges. Progress has been achieved in addressing the crisis and laying the foundations for recovery. Policy steps aimed at supporting growth in the major industrialized countries and important policy actions leading to stronger performance in some emerging markets have improved the economic outlook. A number of substantial challenges still remain. We therefore renew our commitment to pursue appropriate macroeconomic policies and structural reforms. These will contribute to more balanced growth in the world economy, thereby reducing external imbalances.

【図36】

英語閲覧対象文書の単語認定結果を示す図

I. Getting[getting] the World[word] Economy[economy] on Track[track] for Sustained[sustained] Growth[growth]

4. Since we met[met] last year in Birmingham[birmingham], the world[word] economy[economy] has faced[faced] major[major] challenges[challenges]. Progress[progress] has been achieved[achieved] in addressing[addressing] the crisis[crisis] and laying[laying] the foundations[foundations] for recovery[recovery]. Policy[policy] steps[steps] aimed[aimed] at supporting[supporting] growth[growth] in the major[major] industrialized[industrialized] countries[countries] and important policy[policy] actions[actions] leading[leading] to stronger[stronger] performance[performance] in some emerging[emerging] markets[markets] have improved[improved] the economic[economic] outlook[outlook]. A number of substantial[substantial] challenges[challenges] still remain[remain]. We therefore renew[renew] our commitment[commitment] to pursue[pursue] appropriate[appropriate] macroeconomic[macroeconomic] policies[policies] and structural[structural] reforms[reforms]. These will contribute[contribute] to more balanced[balanced] growth[growth] in the world[word] economy[economy], thereby[thereby] reducing[reducing] external[external] imbalances[imbalances].

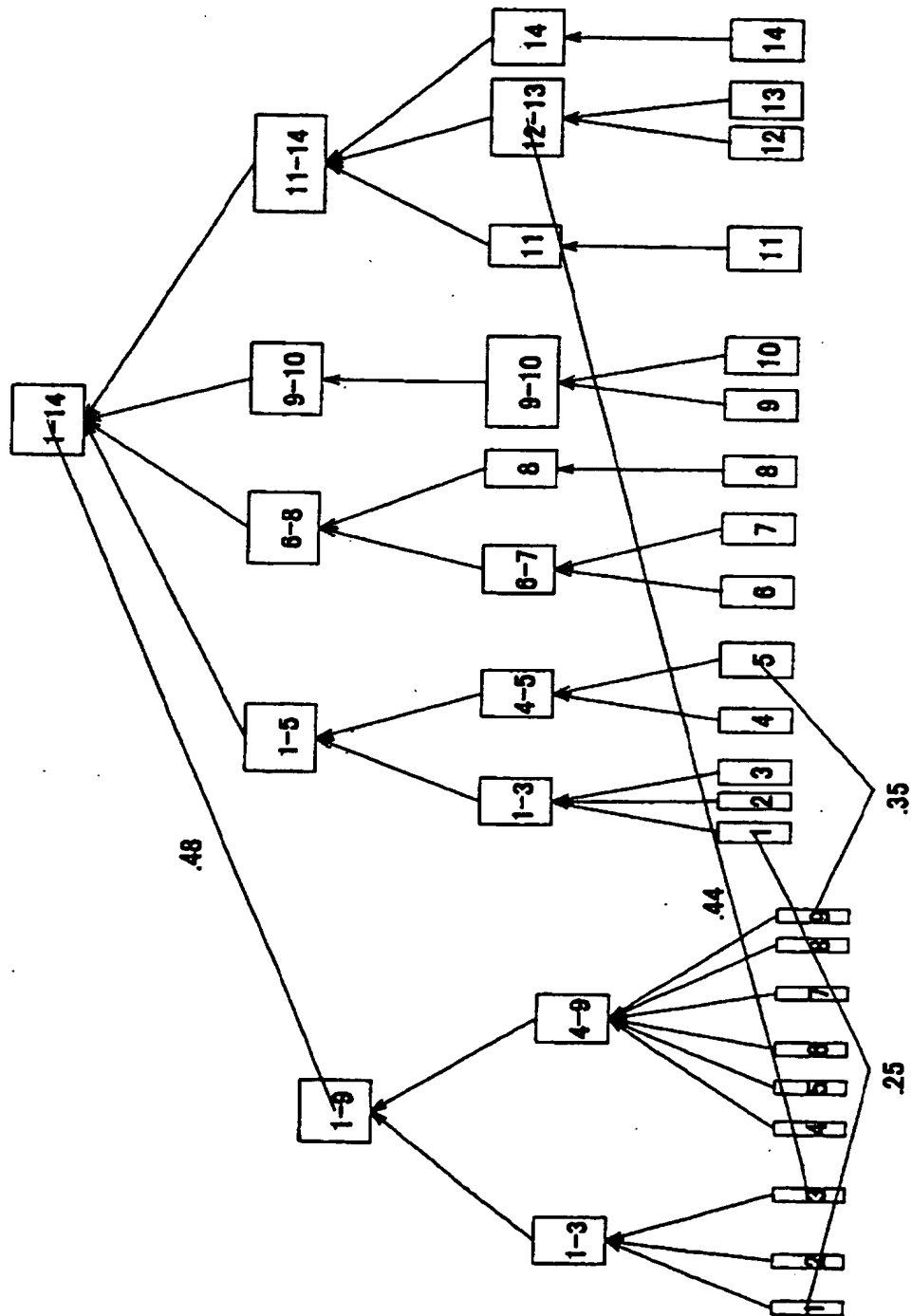
【図37】

ストップワードの例を示す図

about, across, again, against, all, along, already, also, among, an, and, another, are, area, areas, around, as, asked, asking, at, back, be, because, been, before, behind, being, best, better, between, both, but, by, can, certain, clearly, come, different, do, during, each, early, end, enough, even, every, everyone, face, fact, few, first, for, from, fully, further, general, gets, give, given, go, good, goods, greater, groups, has, have, having, high, higher, how, if, important, in, interest, interesting, interests, into, is, it, its, keep, know, last, later, least, less, longer, made, make, many, member, members, more, most, much, must, necessary, need, needed, needs, new, next, no, not, now, number, of, older, on, once, one, only, open, opens, order, other, others, our, out, over, part, place, point, points, possible, present, presented, problem, problems, put, right, same, shall, should, shows, since, small, so, some, states, still, such, take, taken, than, that, the, their, them, then, there, therefore, these, they, this, those, three, through, thus, to, together, too, toward, two, under, up, upon, us, use, was, way, ways, we, well, what, when, where, which, while, who, will, with, within, without, work, working, works, would, year, years, yet, young

【図38】

英語の関連話題の抽出結果を示す図



【図 3 9】

英語の関連箇所の要約例を示す図

(その1)

<p>KÖLN 1999 [1] ... Progress has been achieved in addressing the crisis and laying the foundations for recovery. ... The world economy is still feeling the effects of the financial crises that started in Asia two years ago. ...</p>	<p>OKINAWA 2000 [1] ... Yet the financial and economic crises of the past few years have presented enormous challenges for the world economy. Together with many of our partners around the world, we have devoted ourselves to alleviating the adverse effects of the crisis, stimulating economic recovery, and identifying ways to help prevent future upheavals, including measures to strengthen the international financial architecture. ...</p>
--	---

【図 4 0】

英語の関連箇所の要約例を示す図(その2)

<p>KÖLN 1999 [3]</p> <p>II. Building a World Trading System That Works for Everyone</p> <p>The multilateral trading system incorporated in the World Trade Organization (WTO) has been key to promoting international trade and investment and to increasing economic growth, employment and social progress. ... Given the WTO's vital role, we agree on the importance of improving its transparency to make it more responsive to civil society while preserving its government-to-government nature. ... We therefore call on all nations to launch at the WTO Ministerial Conference in Seattle in December 1999 a new round of broad-based and ambitious negotiations with the aim of achieving substantial and manageable results. ... In this context we reaffirm our commitment made in Birmingham last year to the least developed countries on improved market access. We also urge greater cooperation and policy coherence among international financial, economic, labor and environmental organizations. ...</p>	<p>OKINAWA 2000 [12-13]</p> <p>... The adoption of the short-term package in Geneva, regarding implementation of Uruguay Round undertakings, increased market access for the LDCs, technical assistance for enhanced capacity building) as well as improvement in WTO transparency, was an important first step in this direction and must be pursued expeditiously. ... We must ensure that the multilateral trading system is strengthened and continues to play its vital role in the world economy. ... We agree that the objective of such negotiations should be to enhance market access, develop and strengthen WTO rules and disciplines, support developing countries in achieving economic growth and integration into the global trading system, and ensure that trade and social policies, and trade and environmental policies are compatible and mutually supportive. ... In this regard, international and domestic policy coherence should be enhanced, and co-operation between the international institutions should be improved. ...</p>
---	---

【図 4 1】

英語の関連箇所の要約例を示す図

(その3)

<p>KÖLN 1999 [9]</p> <p>... We also welcome the increasing cooperation between the ILO and the IFIs in promoting adequate social protection and core labor standards. ... In addition, we stress the importance of effective cooperation between the WTO and the ILO on the social dimensions of globalization and trade liberalization.</p>	<p>OKINAWA 2000 [5]</p> <p>... We also welcome the increasing co-operation between the International Labour Organisation (ILO) and the International Financial Institutions (IFIs) in promoting adequate social protection and core labour standards. ... In addition, we stress the importance of effective co-operation between the World Trade Organisation (WTO) and the ILO on the social dimensions of globalisation and trade liberalisation. ...</p>
---	---

【書類名】 要約書

【要約】

【課題】 複数の文書に含まれる粒度の異なる話題について、適切な関連箇所を切り出して提示することが課題である。

【解決手段】 話題階層認定手段 1 は、複数の閲覧対象文書のそれぞれの話題階層を認定し、関連話題抽出手段 2 は、認定された話題階層に基づき、複数の文書に共通して現れる関連話題を抽出する。関連話題提示手段 3 は、複数の文書の比較作業を支援するために、抽出された関連話題に対応する記述箇所を、それぞれの文書から切り出し、関連箇所として出力する。

【選択図】 図 1

出 願 人 履 歴 情 報

識別番号 [000005223]

1. 変更年月日	1996年 3月26日
[変更理由]	住所変更
住 所	神奈川県川崎市中原区上小田中4丁目1番1号
氏 名	富士通株式会社